

APPROXIMATE NUMBER SYSTEM AND MATHEMATICS

EDITED BY: Jingguang Li, Xinlin Zhou and Marcus Lindskog
PUBLISHED IN: Frontiers in Psychology





frontiers

Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88963-208-4

DOI 10.3389/978-2-88963-208-4

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

APPROXIMATE NUMBER SYSTEM AND MATHEMATICS

Topic Editors:

Jingguang Li, Dali University, China

Xinlin Zhou, Beijing Normal University, China

Marcus Lindskog, Uppsala University, Sweden

Humans process quantity information without the aid of language or symbols to guide a variety of everyday life decisions. The cognitive system that supports this intuitive skill is often referred to as the approximate number system (ANS). It has been argued that the ANS serves as the foundation of the formal symbolic number system—mathematics. Abundant empirical evidence is supportive of this view: acuity of the ANS is positively correlated with symbolic math performance, training of the ANS may cause improvements in symbolic math performance, and the ANS and symbolic number processing may share a common neural underpinning. However, recently several theories and empirical data cast doubt on the role of the ANS in symbolic math processing. This e-book aims to advance our understanding of the underlying mechanisms of the overlap between the ANS and mathematics.

Citation: Li, J., Zhou, X., Lindskog, M., eds. (2019). Approximate Number System and Mathematics. Lausanne: Frontiers Media. doi: 10.3389/978-2-88963-208-4

Table of Contents

- 05 Editorial: Approximate Number System and Mathematics**
Jingguang Li, Xinlin Zhou and Marcus Lindskog
- 08 Regular Distribution Inhibits Generic Numerosity Processing**
Wei Liu, Yajun Zhao, Miao Wang and Zhijun Zhang
- 19 On the Difference Between Numerosity Processing and Number Processing**
Anne H. van Hoogmoed and Evelyn H. Kroesbergen
- 32 Using Hierarchical Linear Models to Examine Approximate Number System Acuity: The Role of Trial-Level and Participant-Level Characteristics**
Emily J. Braham, Leanne Elliott and Melissa E. Libertus
- 46 Comparing Numerical Comparison Tasks: A Meta-Analysis of the Variability of the Weber Fraction Relative to the Generation Algorithm**
Mathieu Guillaume and Amandine Van Rinsveld
- 55 The Acuity and Manipulability of the ANS Have Separable Influences on Preschoolers' Symbolic Math Achievement**
Ariel Starr, Rachel C. Tomlinson and Elizabeth M. Brannon
- 63 A Mechanistic Study of the Association Between Symbolic Approximate Arithmetic Performance and Basic Number Magnitude Processing Based on Task Difficulty**
Wei Wei, Wanying Deng, Chen Chen, Jie He, Jike Qin and Yulia Kovas
- 72 The Role of Approximate Number System in Different Mathematics Skills Across Grades**
Dan Cai, Linni Zhang, Yan Li, Wei Wei and George K. Georgiou
- 82 Implications of Change/Stability Patterns in Children's Non-symbolic and Symbolic Magnitude Judgment Abilities Over One Year: A Latent Transition Analysis**
Cindy S. Chew, Jason D. Forte and Robert A. Reeve
- 95 Differences in Counting Skills Between Chinese and German Children are Accompanied by Differences in Processing of Approximate Numerical Magnitude Information**
Jan Lonnemann, Su Li, Pei Zhao, Janosch Linkersdörfer, Sven Lindberg, Marcus Hasselhorn and Song Yan
- 103 Numerical Processing Impairment in 22q11.2 (LCR22-4 to LCR22-5) Microdeletion: A Cognitive-Neuropsychological Case Study**
Livia de Fátima Silva Oliveira, Annelise Júlio-Costa, Fernanda Caroline dos Santos, Maria Raquel Santos Carvalho and Vitor Geraldi Haase
- 124 Approximate Arithmetic Training Improves Informal Math Performance in Low Achieving Preschoolers**
Emily Szkudlarek and Elizabeth M. Brannon
- 136 Testing the Efficacy of Training Basic Numerical Cognition and Transfer Effects to Improvement in Children's Math Ability**
Narae Kim, Selim Jang and Soohyun Cho

148 *Symbolic Number Comparison is not Processed by the Analog Number System: Different Symbolic and Non-symbolic Numerical Distance and Size Effects*

Attila Krajcsi, Gábor Lengyel and Petia Kojouharova

164 *Task Constraints Affect Mapping From Approximate Number System Estimates to Symbolic Numbers*

Dana L. Chesney and Percival G. Matthews



Editorial: Approximate Number System and Mathematics

Jingguang Li^{1*}, Xinlin Zhou² and Marcus Lindskog³

¹ College of Education, Dali University, Dali, China, ² State Key Laboratory of Cognitive Neuroscience and Learning, IDG/McGovern Institute for Brain Research, Beijing Normal University, Beijing, China, ³ Department of Psychology, Uppsala University, Uppsala, Sweden

Keywords: approximate number system, number sense, non-symbolic number acuity, numerical cognition, mathematics

Editorial on the Research Topic

Approximate Number System and Mathematics

Humans process quantity information without the aid of language or symbols to guide a variety of everyday life decisions. The cognitive system that supports this intuitive skill is often referred to as the approximate number system (ANS). It has been argued that the ANS serves as the foundation of the formal symbolic number system—mathematics (Dehaene, 1997). Abundant empirical evidence is supportive of this view: acuity of the ANS is positively correlated with symbolic math performance (Chen and Li, 2014), training of the ANS may cause improvements in symbolic math performance (Bugden et al., 2016), and the ANS and symbolic number processing may share a common neural underpinning (Piazza et al., 2004). However, recently several theories and empirical data cast doubt on the role of the ANS in symbolic math processing (Reynvoet and Sasanguie, 2016; Leibovich et al., 2017). This Research Topic aims to advance our understanding of the underlying mechanisms of the overlap between the ANS and mathematics.

The first portion of this Research Topic centers on the measurement issue of the ANS. Liu et al. demonstrated that regularity of visual features in the non-symbolic numerical task influenced processing of numerical information. For regular patterns of dot arrays, numerosity processing is inhibited; but for random patterns, numerosity information could be extracted independently of visual features. Thus, to measure ANS acuity, it is necessary to avoid regular dot patterns in the non-symbolic numerical task. van Hoogmoed and Kroesbergen suggested that convex hull, the smallest convex polygon that contains an array of dots, could be a plausible confounding factor in the non-symbolic numerical task. By using event-related potentials (ERP) from electroencephalography recordings, they found no signs of a distance effect for numerosity, but a distance effect for convex hull instead. Consequently, non-numerical visual features might at least partly influence performance in non-symbolic numerical tasks. Hence, it is unclear whether non-numerical visual processing or numerical processing in the non-symbolic numerical task contributes to the widely reported association between ANS acuity and math performance. Furthermore, their ERP data indicated that symbolic and non-symbolic numerosities were processed differentially, questioning if non-symbolic and symbolic numerosities share the same neural circuitry, as previously suggested (e.g., Dehaene, 1997). Braham et al. addressed this issue by using hierarchical linear modeling, which has the advantage of being able to isolate the numerical and non-numerical visual component in non-symbolic numerical task performance both within and between individuals. Critically, they found that only the numerical component contributed to adults' math ability. Finally, Guillaume and Van Rinsveld performed a meta-analysis regarding the variability of the Weber fraction in different versions of the non-symbolic number comparison paradigm. They found that different methods used for controlling for non-numerical information cause highly variable Weber fraction scores. Accordingly, they recommended not to compare Weber fraction scores from different tasks.

OPEN ACCESS

Edited and reviewed by:

Bernhard Hommel,
Leiden University, Netherlands

*Correspondence:

Jingguang Li
jingguang.li.k@gmail.com

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 24 August 2019

Accepted: 27 August 2019

Published: 12 September 2019

Citation:

Li J, Zhou X and Lindskog M (2019)
Editorial: Approximate Number
System and Mathematics.
Front. Psychol. 10:2084.
doi: 10.3389/fpsyg.2019.02084

The second portion of this Research Topic focuses on the correlation between ANS acuity and math ability. Testing this correlation is the first step for further investigation of the causal relationship between ANS and math performance. Starr et al. suggested a new path underlying the association between ANS and math performance. They found that ANS manipulability (i.e., the ability to perform arithmetic operations on approximate numerical quantities) positively predicted math achievement in preschool children, and the predictive power of ANS manipulability was independent of the influence of ANS acuity. Wei et al. examined the relationship between number magnitude processing and symbolic approximate arithmetic performance (i.e., the ability to provide an approximate answer to an arithmetic question), which should arguably be largely uninfluenced by language. They found that both semantic and spatial number processing (indexed by the two-digit number comparison and number-line estimation task, respectively) are positively correlated to the symbolic approximate arithmetic performance, and these associations are moderated by the task difficulty of the symbolic approximate arithmetic task.

Two studies demonstrated that the correlation between ANS acuity and math performance is moderated by multiple factors. Cai et al. found that the correlation between ANS acuity and math performance varies across different grade levels (kindergarten vs. primary school), type of math tests, and type of ANS tests (non-symbolic estimation vs. number-line task). Using latent class modeling, Chew et al. identified four different magnitude ability profiles based on children's performance in the non-symbolic and symbolic numerical task. Further, they observed both stability and change in the four different profiles across a 1-year time period. Finally, profile membership was differentially related math performance at different ages.

Another two studies revealed that differences in math ability of different populations could be attributed to differences in ANS acuity. Lonnemann et al. found that Chinese children have better counting skills than their German peers. More importantly, the advantages in counting in Chinese children were accompanied by superior performance in a non-symbolic numerical comparison task. In addition, Oliveira et al. reported a case study on a girl with specific numerical processing impairment and a rare genetic disorder—22q11.2 deletion syndrome. The girl has normal general intelligence; however, she manifested severe deficits in single-digit calculation accompanied by poor performance in the non-symbolic numerical comparison task.

The third portion of this Research Topic examines whether training of the ANS leads to improvement in symbolic math performance. The training approach not only tests the causal relationship between ANS acuity and math performance, but also provides valuable insights for math education (Bugden et al., 2016). Szkudlarek and Brannon found a transfer effect from ANS training to math performance. A group of preschool children trained for

1 month with a computer-based non-symbolic arithmetic training program. After controlling for confounding factors, children with low math abilities in the ANS-training group outperformed control-group children on informal symbolic math problems. In contrast, Kim et al. did not find a transfer effect in their training experiment with first-grade children. Although significant improvement in ANS acuity was observed following a 6-week training period, children showed no improvement in math performance. To resolve the discrepancies between the above two training studies, more replication studies with rigorous methodologies are needed (Szucs and Myers, 2017).

The final portion of this Research Topic examines the distinction and mapping between the ANS and the symbolic numerical processing system by analyzing psychophysical features of different non-symbolic and symbolic numerical tasks. Krajcsi et al. made an extensive comparison of the several psychophysical properties of non-symbolic and symbolic number comparison, including error rates, reaction times, and diffusion-model drift rates. They found that the ratio-based ANS model only fits the non-symbolic number comparison data, but not the symbolic comparison data. Accordingly, the authors argued that different cognitive systems are in charge of symbolic and non-symbolic number processing. Chesney and Matthews found that different versions of non-symbolic numerosity tasks give rise to differences in performance. More specifically, while a free estimation task showed a classical pattern of scalar variability there was no evidence for this error pattern in a number-line and ratio estimation task. Furthermore, participants showed underestimation in the free estimation task but accurate estimation in the ratio task. They argued that these task constraints affect the ANS-math mapping process.

Taken together, this Research Topic combines diverse methodologies to advance our understanding of the relationship between the approximate number system and mathematics. According to the new data in this Research Topic, it might be too simple to conclude that the ANS and math are related or separated. Instead, it is worth asking how (i.e., the cognitive paths) and when (i.e., different developmental stages, task variants, and types of participants) the ANS is linked to math.

AUTHOR CONTRIBUTIONS

JL wrote the first draft. JL, XZ, and ML contributed to the revision of the paper.

FUNDING

The organization of this Research Topic was supported by the National Natural Science Foundation of China (31500884) and the Innovation Team of Dali University (SKPY2019303) to JL and by a grant from The Swedish Foundation for Humanities and Social Sciences (P15-0430:1) to ML.

REFERENCES

- Bugden, S., DeWind, N. K., and Brannon, E. M. (2016). Using cognitive training studies to unravel the mechanisms by which the approximate number system supports symbolic math ability. *Curr. Opin. Behav. Sci.* 10, 73–80. doi: 10.1016/j.cobeha.2016.05.002
- Chen, Q., and Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychol.* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. New York, NY: Oxford University Press.
- Leibovich, T., Katzin, N., Harel, M., and Henik, A. (2017). From “sense of number” to “sense of magnitude”: the role of continuous magnitudes in numerical cognition. *Behav. Brain Sci.* 40:E164. doi: 10.1017/S0140525X16000960
- Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* 44, 547–555. doi: 10.1016/j.neuron.2004.10.014
- Reynvoet, B., and Sasanguie, D. (2016). The symbol grounding problem revisited: a thorough evaluation of the ANS mapping account and the proposal of an alternative account based on symbol–symbol associations. *Front. Psychol.* 7:1581. doi: 10.3389/fpsyg.2016.01581
- Szucs, D., and Myers, T. (2017). A critical analysis of design, facts, bias and inference in the approximate number system training literature: a systematic review. *Trends Neurosci. Educ.* 6, 187–203. doi: 10.1016/j.tine.2016.11.002

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Li, Zhou and Lindskog. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Regular Distribution Inhibits Generic Numerosity Processing

Wei Liu¹, Yajun Zhao², Miao Wang¹ and Zhijun Zhang^{3*}

¹ School of Education, Yunnan Minzu University, Kunming, China, ² School of Sociology and Psychology, Southwest University for Nationalities, Chengdu, China, ³ Department of Psychology and Behavioural Sciences, Zhejiang University, Hangzhou, China

This study investigated the role of pattern regularity in approximate numerical processing. Experiment 1 demonstrated that the change in stimulus size has a distinct effect on the adaptation aftereffect for random and regular patterns. For regular patterns, adapting to large patterns and being tested with small patterns caused stronger aftereffects than the reverse treatment, in which the participants adapted to small patterns and were tested with large patterns. For random patterns, this effect was absent. Experiment 2 revealed a distinct connectedness effect on the numerosity processing of random and regular patterns. For random patterns, reference stimuli were perceived to contain fewer items when the dots were connected by lines than when they were not connected, and the number of items in the connected reference was further underestimated when the participants adapted to unconnected patterns with the same number of dots. For regular patterns, this effect was absent. Distinct mechanisms were thus suggested for the numerosity coding of random and regular patterns. For random patterns, the change in primary texture features would be abstracted from numerosity processing, while connectedness could affect this coding by affecting the processing of numerical unit individuation. For regular patterns, generic numerosity processing is inhibited, and numerical judgments appear to be inferred from the visual processing results of texture features such as dot size or the distance between adjacent dots.

Keywords: numerosity perception, element distribution, connectedness effect, individuation, texture specificity

OPEN ACCESS

Edited by:

Xinlin Zhou,
Beijing Normal University, China

Reviewed by:

Robert Reeve,
The University of Melbourne, Australia
Dawei Li,
Duke University, United States

*Correspondence:

Zhijun Zhang
zjzhang@zju.edu.cn

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 13 April 2018

Accepted: 09 October 2018

Published: 31 October 2018

Citation:

Liu W, Zhao Y, Wang M and
Zhang Z (2018) Regular Distribution
Inhibits Generic Numerosity
Processing. *Front. Psychol.* 9:2080.
doi: 10.3389/fpsyg.2018.02080

INTRODUCTION

Numerosity cognition is accompanied by the processing of a combination of visual features (Dehaene, 1992; Franconeri et al., 2009). Previous studies have suggested the independence of numerosity processing from the processes associated with texture features, and the abstraction process is suggested to be part of numerosity coding (Burr and Ross, 2008a,b; Liu et al., 2012, 2013). However, these studies have been challenged by other studies indicating that perceived numerosity is affected by some visual features, such as size, contrast, and density (Dakin et al., 2011; Raphael et al., 2013; Raphael and Morgan, 2015). Numerosity adaptation, in which the numerosity of the adaptor affects the observer's perception of quantity, can be inferred from the change in perceived numerosity before and after adaptation (adaptation aftereffect). Numerosity adaptation is proposed as evidence of an independent numerosity processing mechanism. However, other researchers have argued that this adaptation could occur via more general texture-like mechanisms, relying on features such as dot size or texture density adaptation (Durgin, 2008).

The interaction between visual properties and numerosity coding seems to contradict the idea that numerosity processing occurs through an independent mechanism. Numerosity processing has been proposed to consist of several steps that involve distinct levels (Liu et al., 2013; Zhang et al., 2014). One way to explain the mentioned contradiction is to analyze the interaction at a specific level of numerical processing.

Numerosity processing begins with primary texture analyses. The combined computation results for surrogate features, such as size, density, and the average distance between adjacent dots, are first processed by the visual system. Common bases in processing suggest a pathway between numerosity processing and texture processing, and the interaction may occur mainly at the primary level. In a study by Anobile et al. (2014), participants were asked to compare the number or density of pairs of dot arrays, and the Weber fraction was analyzed. With moderate density, the thresholds increased with numerosity. When the dots became denser, a new pattern of change appeared, suggesting a density-processing mode, regardless of whether participants compared the numerosity or the density of stimulus pairs. As the dots became denser, it became difficult to separate individual dots as numerical units within the crowded texture. Under that condition, numerical cognition was inhibited, and density cognition superseded this processing (Anobile et al., 2014, 2015). When numerical coding is inhibited, stimulus processing may consist of no more than texture processing, which is frequently affected by visual features (Liu et al., 2017).

As visual information is processed from primary to higher levels along the ventral pathway, the presentation of the information transforms from a specific to an abstract format (Dehaene and Changeux, 1993), and the underlying neuronal bases shift from simple to complex (Liu et al., 2017). Generic numerosity processing involves the function of high-level processes such as individuation, abstraction, and numerical unit representation (Liu et al., 2017). The existence of individuation in numerosity processing can be demonstrated by the connectedness effect. When randomly distributed dots are connected by lines, the perceived magnitude is significantly reduced. Two connected dots are considered to be one when observers compare the number of dots (Franconeri et al., 2009; He et al., 2009, 2015; Milne et al., 2013). Adaptation causes a further reduction in the estimated numerosity of connected dots. In a study by Fornaciai et al. (2016), adaptation to a 20-dot pattern (the same number of dots as in the reference) caused a further reduction in the estimated numerosity of the reference, in which two dots were connected as one pair. This fact suggests that adaptation to numerosity acts on perceived numerosity and that magnitude estimation is based on the individuation of items.

The individuation and presentation of numerical units are necessary in numerosity processing (Gallistel and Gelman, 2000). The inhibition of individuation is synchronic with the inhibition of generic numerosity processing (Liu et al., 2013, 2017). A crowding-like effect may inhibit numerosity processing because dots are too dense to be individuated (Anobile et al., 2014). A high degree of regularity in the distribution of dots (e.g., dots spaced at a uniform distance or aligned in rows) could be another way to inhibit numerosity processing because

dots in such a distribution are also difficult to individuate. The overall configuration emphasizes meaningful information and observers are likely to understand the pattern by analyzing the spatial relationships between one dot and its fellows in another “neighborhood” instead of by separating a single dot and analyzing it without context (Liu et al., 2017). A distinct adaptation aftereffect was revealed in the numerosity processing of randomly and regularly distributed dots, suggesting this inhibition in the coding of regular dots. The numerosity adaptation aftereffect was immune to change in the orientation of the elements between adaptors and tests and, furthermore, showed binocular transfer (Durgin, 2001; Harris et al., 2011; Sweeny et al., 2011) in the coding of randomly distributed patterns. However, the adaptation aftereffect was specific to the change in the orientation of the elements and exhibited monocular transfer in the coding of regularly distributed patterns (Liu et al., 2017). Numerosity processing should not be generic based on the visual coding of regular patterns.

Texture coding typically interacts at the primary processing level, whereas individuation involves higher levels of activity (Liu et al., 2017). If the distinguishable processes exclusively pertaining to numeral coding, such as individuation or abstraction, are what determine the independence of numerosity cognition, then the arguments claiming that various visual features affect numerosity processing would not necessarily be contradictory. In our 2017 study, it was proposed that element orientation has distinct effects on numerosity processing in random and regular patterns and that compared with random dots, regularly distributed dots inhibit high-level numeral processing. Evidence showing dissociation in numerosity adaptation between the coding of random and regular patterns would convincingly support the case that generic numerosity coding is independent of texture coding and that numerosity coding interacts with texture coding when certain processes are inhibited. In the current study, converging evidence was collected to support the hypothesis that distinct mechanisms control the coding of random and regular patterns. Moreover, we provided further evidence supporting the case that numerosity processing of regular patterns depends on analyses of surrogate features and that perceived numerosity can be inferred from the processing results of certain features of visual arrays, such as dot size or distance. In addition, we collected clearer evidence suggesting that individuation was inhibited in the numerosity processing of regular patterns.

Two experiments were conducted using the adapting paradigm (Burr and Ross, 2008a; Fornaciai et al., 2016). If generic numerosity processing is inhibited when regularly distributed patterns are coded, then it is possible that numerosity estimation is inferred via texture-like mechanisms, such as estimation of the size of the dots or the distance between them (Sophian, 2007). Therefore, the element size relationship between adaptors and test stimuli could affect the numerosity adaptation aftereffect for regularly distributed patterns (Experiment 1), although size immunity has been confirmed in numerosity adaptation for randomly distributed patterns (Burr and Ross, 2008a; Liu et al., 2012). In addition, it has been proposed that individuation is inhibited with regard to the numerosity

estimation of regular patterns (Liu et al., 2017). We made further efforts to investigate whether individuation is inhibited with regard to numerosity adaptation. We proposed that when the number of dots is equal in a regularly distributed reference and adaptor, even if the dots in the reference are connected, the adaptors will cause no reduction in the estimated magnitude of the reference (Experiment 2), although such a reduction has been revealed for randomly distributed dots (Fornaciai et al., 2016).

EXPERIMENT 1: THE ELEMENT SIZE SPECIFICITY OF NUMEROSITY ADAPTATION WITH RANDOM AND REGULAR DOTS

Experiment 1 investigated whether adaptors with elements whose size was different from those in tests would affect the adaptation aftereffect and whether such an effect would vary between adaptation to random and regular patterns. A paradigm similar to that in the previous study (Burr and Ross, 2008a) was adopted to investigate the numerosity adaptation aftereffect.

Methods Statement

For all experiments, all administered measures and tested experimental conditions were reported. All recorded data from the participants were included in the calculation. Missing data (responses after 1,000 ms in the response window) were excluded from the total set of responses when the selection probability for the point of subjective equality (PSE) was calculated. For each participant, the missing data amounted to less than 3%.

Ethics Statement

The data in Experiments 1 and 2 were analyzed anonymously. All adults in this study's experiments provided their informed consent in both verbal and written forms, and they were compensated for their participation. The ethics committee of Yunnan Minzu University approved this study.

Participants

The sample sizes in our previously published study with a similar paradigm (Liu et al., 2017) were taken into consideration. We collected data from 16 participants in each experiment because the abovementioned study showed that this sample size yields ample power. The participants had either normal or corrected-to-normal vision, and they were right-handed. Six males and 10 females (age range = 19–32 years) participated in Experiment 1.

Apparatus

The stimuli were displayed using E-Prime 1.0 on a 17" monitor (Philips, flat-screen) with a resolution of 1,024 × 768 pixels and a refresh rate of 85 Hz. The experiments were conducted in a dark room, and the viewing distance was approximately 55 cm.

Stimuli

Stimuli were generated using Walk Script 1.0 (ZJU Walkinfo Co., Ltd., Hangzhou, China). During the experiment, stimulus

patterns were all presented within two-fixed circles in the middle of the computer screen (**Figure 1**). Each grayscale pattern (RGB: 128, 128, 128) had a diameter of 300 pixels and was presented against a dark-gray (RGB: 120, 120, 120) background. In the adaptation stage, the two circles served to display the adaptors; in the testing stage, they served to display the reference and test stimuli.

For adaptors, there were 68 rectangular dots presented in one circle and 8 in the other (**Figure 1**). In each circle, half of the dots were white, and the other half were black. The dots in the adaptors were randomly distributed in the "random" condition and were classified into vertical queues by color in the "regular" condition. Each dot in the adaptors was 6 × 6 pixels in the "small" condition and 14 × 14 pixels in the "large" condition. Note that no more than 68 dots were assigned in the adaptors because increasing density would cause increasing difficulty in separating numerical units when more dots were included in the circle, especially, when the "large" adaptors were presented.

Each reference contained 40 dots, which were similar to those in the adaptors. In other words, there were references in which the dots were of small or large size, placed in either random or regular spatial distributions. Within each treatment, the distribution of the dots was kept constant between adaptors and references (random-random or regular-regular), while the dot size differed (small-large or large-small).

For tests, the size and distribution of the dots were kept identical to their references, while the numbers of dots varied. An equidistant logarithmic scale was adopted to decide the numbers of test dots (Dehaene et al., 2008). Moreover, we chose numbers with which a symmetric pattern could be constructed in regular groups; thus, the tests contained 24, 30, 33, 36, 40, 44, 49, 58, or 68 dots. The reference number (40) was assigned in the center of the testing series.

Notably, there was only one distribution pattern for each "regular" stimulus with a certain number of dots (m columns and n rows). Therefore, we also adopted only one picture for each random stimulus, such that equivalent familiarity could be induced for random and regular conditions when the participants performed the experiment. In total, 4 adaptors, 4 references, and 36 test patterns were generated in Experiment 1.

Procedure

We adopted a 2 (dot distribution pattern: random/regular) × 2 (dot size relationship: large-small/small-large) within-subjects design. Therefore, the participants compared the numbers of dots after adaptation across four treatments. Moreover, four unadapted pretests (small-random, small-regular, large-random, and large-regular dots) were conducted as baselines, in which the participants performed the testing procedure directly without any adaptation.

The treatment with adaptation is described in **Figure 2** (random, small-large condition). In each treatment, the participants initiated the first trial by pressing the space bar, and a background frame with two circles and a fixation point was visible during the entire procedure. In the adaptation stage, the background frame lasted for 200 ms. Then, the adaptors were presented in the circles for 1,000 ms.

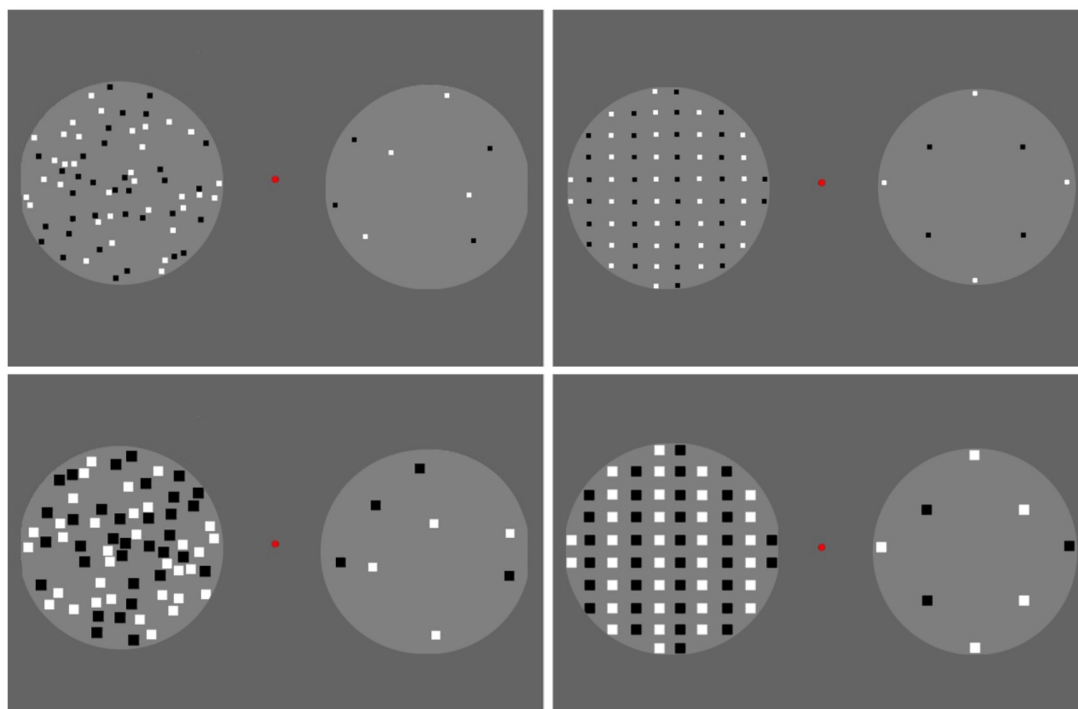


FIGURE 1 | Adaptors used in Experiment 1. The randomly distributed adaptors are shown on the left-hand side, and the regularly distributed (vertical queues) adaptors are shown on the right-hand side. The small adaptors are shown in the upper row, and the large adaptors are shown in the lower row. Dots were generated within two-fixed circles. A total of 68 dots were displayed in one circle, and eight dots were arranged in the other.

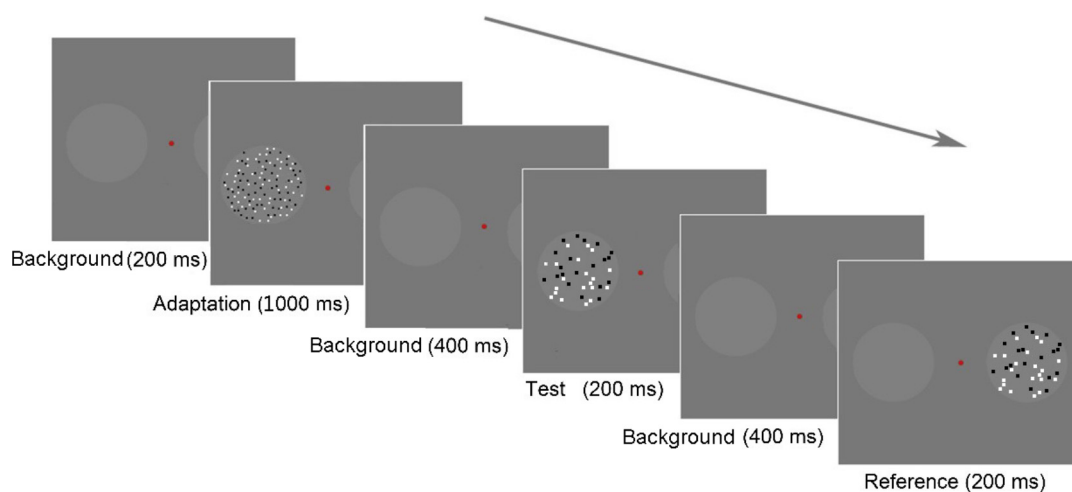


FIGURE 2 | Schematic of the paradigm in Experiment 1. Each trial began with a background frame for 200 ms, followed by an adaptation stage that lasted for 1,000 ms. The test stage began with a background lasting for 400 ms. A test stimulus was displayed in the left circle for 200 ms, followed by a background for 400 ms, and then the reference stimulus was displayed in the right circle for 200 ms. The participants were asked to report, which circle appeared to contain more dots by pressing the appropriate key; if they were uncertain, they were required to guess.

In the testing stage, the background frame was shown for 400 ms at the beginning. Subsequently, a test stimulus was presented in the left circle for 200 ms, followed by the background frame for 400 ms. Then, a reference stimulus was presented in the right circle for 200 ms. Once the reference appeared, the

participants should respond to a forced-choice question: “Which circle contained more dots?” They pressed either the “f” key on the keyboard with their left hand, indicating that the left circle contained more dots, or “j” with their right hand, indicating that the right circle contained more dots. In other words, we used a

two-alternative forced-choice (2AFC) task to assess numerosity perception. The next trial began either after the participant's response or after 1,200 ms without a response.

At the beginning of Experiment 1, brief practice trials with feedback were conducted to improve the participants' familiarity with the formal experiment. Then, the participants completed four pretests in a random sequence before performing the adaptation tasks to create baselines. In the pretest, no adaptors were presented, and each of the 72 trials proceeded directly to the testing stage. After these tasks, the participants began the formal experiment with four treatments, each with 72 trials. The adaptors, reference, and test positions were counterbalanced across participants and were kept identical within treatments for each participant. The sequences of treatments with adaptors were also counterbalanced across participants. Sufficient rest was provided between treatments to avoid fatigue.

Results

Cumulative normal models were fitted to the psychometric functions of each participant using the psignifit toolbox version 2.5.41 for MATLAB¹. The maximum likelihood method (Wichmann and Hill, 2001) was adopted to measure the

magnitude of the connectedness effect. The values of the test stimuli (X -axis) corresponding to the 50% points were calculated from the fitted curves (Figure 3). These values were the PSEs representing the number of test dots that appeared to be equal to the number of reference dots according to each participant. The change in numerosity perception in the tests is represented by the difference in the PSEs under different circumstances (Table 1). Therefore, the magnitude of the numerosity adaptation aftereffect is revealed by the PSEs under the adaptation conditions minus the PSEs in the pretests.

No significant main effect or interaction was observed between the four pretests. There was a significant difference between the treatment and its baseline (pretest) in the random group for the "large-small" condition, $t(15) = 3.09$, $p = 0.008$, $d = 0.77$, and for the "small-large" condition, $t(15) = 3.48$, $p = 0.003$, $d = 0.87$, as well as a significant difference in the regular group for the "large-small" condition, $t(15) = 6.69$, $p < 0.001$, $d = 1.67$, showing that both the randomly and regularly distributed adaptors affected the participants' numerosity perception. In most cases, when the presented scene was shifted from adaptors to tests, the number of dots in the circle decreased from 68 to less than 68 (according to the test dot number), decreasing the numbers perceived by the participants in the tests (Burr and Ross, 2008a). Subsequently, the apparent number of dots in the reference

¹<http://www.bootstrap-software.com/psignifit/>

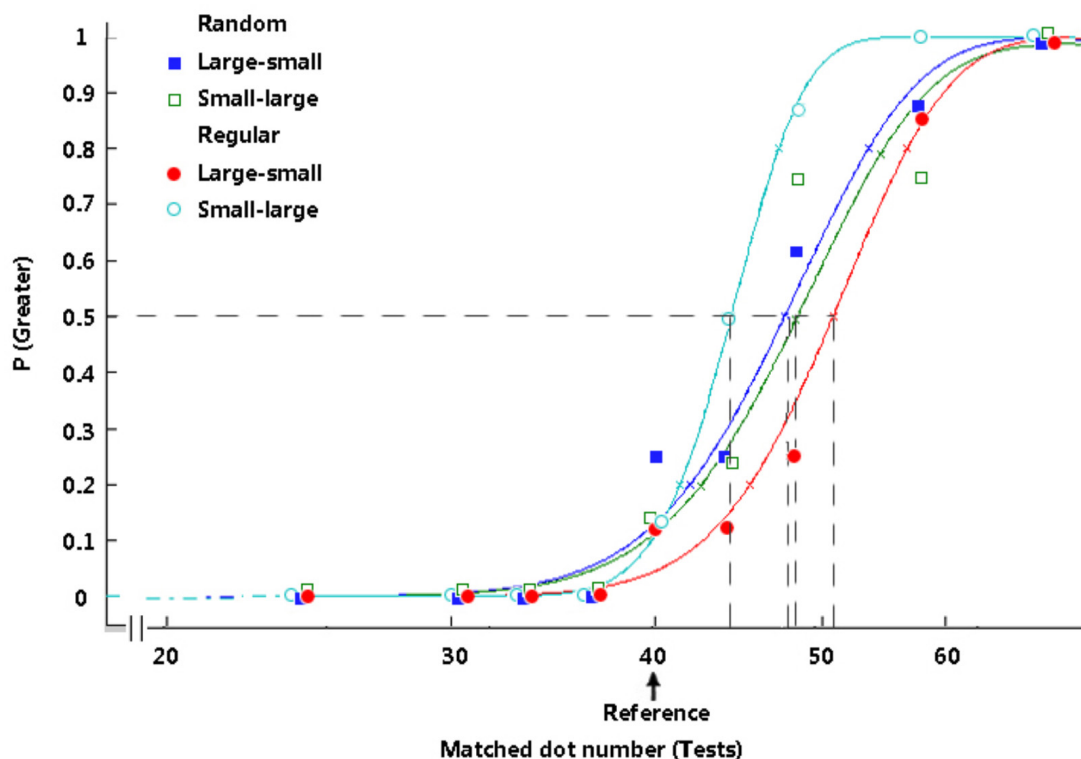


FIGURE 3 | Typical psychometric functions under distinct conditions in Experiment 1. The proportion of trials in which the test stimuli appeared to be more numerous is plotted as a function of the number of test dots, and the vertical dashed lines reveal the PSEs. The arrow indicates the reference number. The participants' typical responding curves are displayed to indicate the average PSE results. In the random group, filled rectangles, dark-blue curve = large adaptors and small tests; open rectangles, green curve = small adaptors and large tests. In the regular group, filled circles, red curve = large adaptors and small tests; open circles, light-blue curve = small adaptors and large tests.

TABLE 1 | The means and SDs for the PSEs in the pretest and adaptation conditions in Experiment 1.

	Test pattern: random				Test pattern: regular			
	Pretest (small)	Pretest (large)	Large–small	Small–large	Pretest (small)	Pretest (large)	Large–small	Small–large
PSE	41.97	42.64	48.03	48.89	41.58	42.40	51.01	44.52
SD	4.03	6.98	7.83	9.81	5.64	4.35	7.59	7.17

PSE, point of subjective equality; SD represents the standard deviation of the PSE. “Large–small” refers to the treatment in which the participants were exposed to large adaptors and small tests. The rest can be performed in the same manner.

(PSE) was overestimated. No significant difference was evident between the treatment and its pretest in the regular group for the “small–large” condition ($p = 0.151$).

The adaptation aftereffect was calculated by subtracting the PSEs of treatments from those of their pretests (Liu et al., 2012, 2017). A 2×2 repeated-measures ANOVA was conducted with the test patterns (random or regular) and the stimulus size relationship (large–small or small–large) as the independent variables and the adaptation aftereffect as the dependent variable. No significant main effect of the test pattern was found ($p = 0.810$); however, the main effect of the stimulus size relationship, $F(1, 15) = 14.38$, $p = 0.002$, $\eta_p^2 = 0.49$, and the interaction between the two factors (Figure 4), $F(1, 15) = 9.93$, $p = 0.007$, $\eta_p^2 = 0.40$, were significant. With regular adaptors and tests, a greater effect was found in the participants’ numerosity perception when they adapted to the large dots and were tested using small dots than when they adapted to small dots and were tested using large dots, $p < 0.001$. When the participants adapted to random dots and their perception was tested using random dots, the size relationship between the adapting and testing

stimuli caused no significant difference between treatments, $p = 0.908$.

EXPERIMENT 2: THE EFFECT OF CONNECTEDNESS ON NUMEROSITY ADAPTATION WITH RANDOM AND REGULAR DOTS

Experiment 2 examined the effects of the connectedness of elements on numerosity adaptation, in which the adapting and testing stages were conducted with randomly and regularly distributed dots, respectively. The aftereffect of numerosity adaptation in connected random dots was tested by Fornaciai et al. (2016). A similar paradigm was used in Experiment 2.

Methods

Participants

Six males and 10 females (age range = 20–32 years) participated in Experiment 2.

Stimuli

Both the reference and the test patterns were arranged within two-fixed circles similar to those in Experiment 1. Four reference patterns were first created, each containing 40 circular dots with a diameter of 12 pixels (Figure 5). In two of the patterns, no lines were included. In the other two patterns, each pattern contained 10 two-pixel-wide line segments of varying length (30–50 pixels). The dots were at least 10 pixels apart. The lines did not cross each other. In Reference 1 (random, connected), the dots were randomly distributed. In each pattern, each individual line linked two adjacent dots to form a connected object (10 lines connected to 20 dots overall). In Reference 2 (random unconnected), the dots were randomly distributed, and no lines were included. In Reference 3 (regular, connected), the dots were arranged into vertical queues. Ten vertical lines were arranged to connect adjacent dots. In Reference 4 (regular, unconnected), the dot presentation was similar to that in Reference 3, and no lines were included. The lines in each random pattern had a varying length, with an average value of 42 pixels, and the lines in each regular pattern had a fixed length of 44 pixels. The connected reference was used for the treatment conditions, and the unconnected reference was used for the baseline conditions. In each condition, the dot distribution in the test patterns was similar to that in the reference patterns (random or regular). No lines were included in the test patterns. The tests contained 18, 24, 30, 33, 36, 40, 44, 49, 58, 68, or 78 dots. Overall, 22 test patterns were generated.

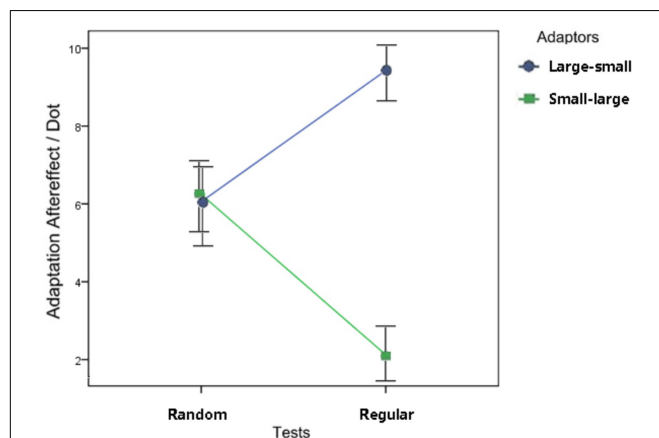


FIGURE 4 | Results of the ANOVA in Experiment 1. A significant interaction was found between the dot distribution (the two shapes on the left of Figure 4 = random adaptors, references, and tests; the other two shapes on the right = regular adaptors, references, and tests) and the size relationship (circles = adapting to large dots and tested by small dots; rectangles = adapting to small dots, and tested by large dots). In the regular groups, a greater adaptation effect was revealed in the large–small condition than in the small–large condition. In the random groups, however, the difference between conditions with those two size relationships was not significant. Error bars represent 1 standard error of the mean.

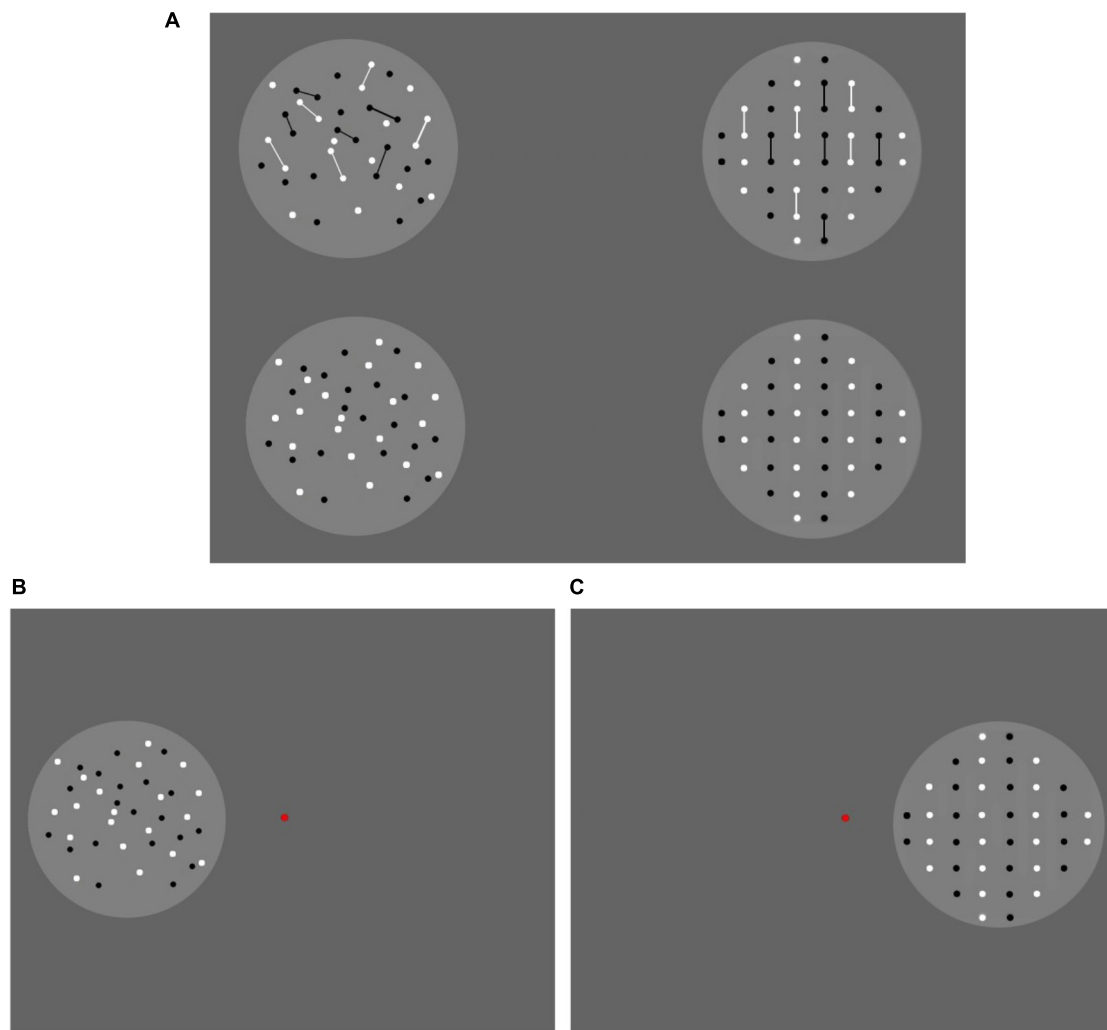


FIGURE 5 | The stimuli used in Experiment 2. The reference patterns are shown in **(A)**. The connected adaptors are displayed in the upper row, and the unconnected adaptors are presented in the lower row. In each row, from left to right, the references in the random and regular groups, respectively. The test patterns were similar to the reference in the lower row, except for the number of dots. The adaptor patterns are shown in **(B)** (the random group) and **(C)** (the regular group). In the adaptation stage, the 40-dot adaptor was presented on one side of the screen. The other side was blank.

For adaptors, the 40-dot test patterns were adopted. In the random group, 40 dots were randomly distributed in the presentation circle. In the regular group, 40 dots were regularly distributed.

Procedure

We adopted a 2 (stimulus pattern: randomly/regularly distributed elements) \times 2 (reference pattern: connected/unconnected dots) within-subjects design. The procedure is described in **Figure 6**. In general, the procedure was similar to that in Experiment 1. There is one observable difference in the testing stage. In this stage, the reference was presented first in the same position where the adaptor was presented before, followed by the test displayed in the opposite position. The participants were asked to compare the number of dots in the reference and test stimuli, that is, to report which circle contained more dots by pressing “f” or “j”;

and they were instructed to ignore the lines (if any) when they were estimating the number of dots.

Results

Figure 7 and **Table 2** demonstrate the difference in the average PSEs under different circumstances. The magnitude of the connectedness effect is indicated by the baseline PSEs minus the treatment PSEs for each group.

There was no significant difference between the PSE of the baselines (the unconnected and unadapted conditions in the random and regular groups) and the standard value (40), $p > 0.05$. In the random group, when the dots were not connected by lines, no significant PSE difference was revealed between the conditions with and without adaptation ($p = 0.247$). Adapting to an adaptor with an equal reference number did not affect the participants’ numerosity perception of the reference.

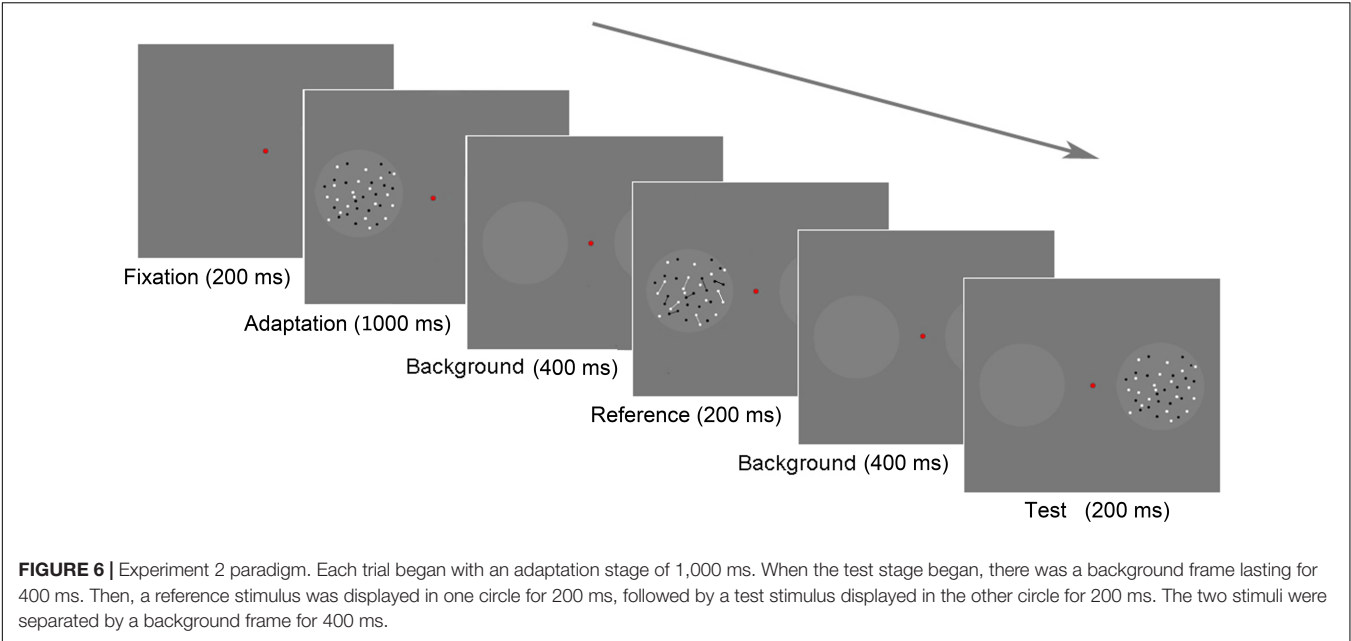


FIGURE 6 | Experiment 2 paradigm. Each trial began with an adaptation stage of 1,000 ms. When the test stage began, there was a background frame lasting for 400 ms. Then, a reference stimulus was displayed in one circle for 200 ms, followed by a test stimulus displayed in the other circle for 200 ms. The two stimuli were separated by a background frame for 400 ms.

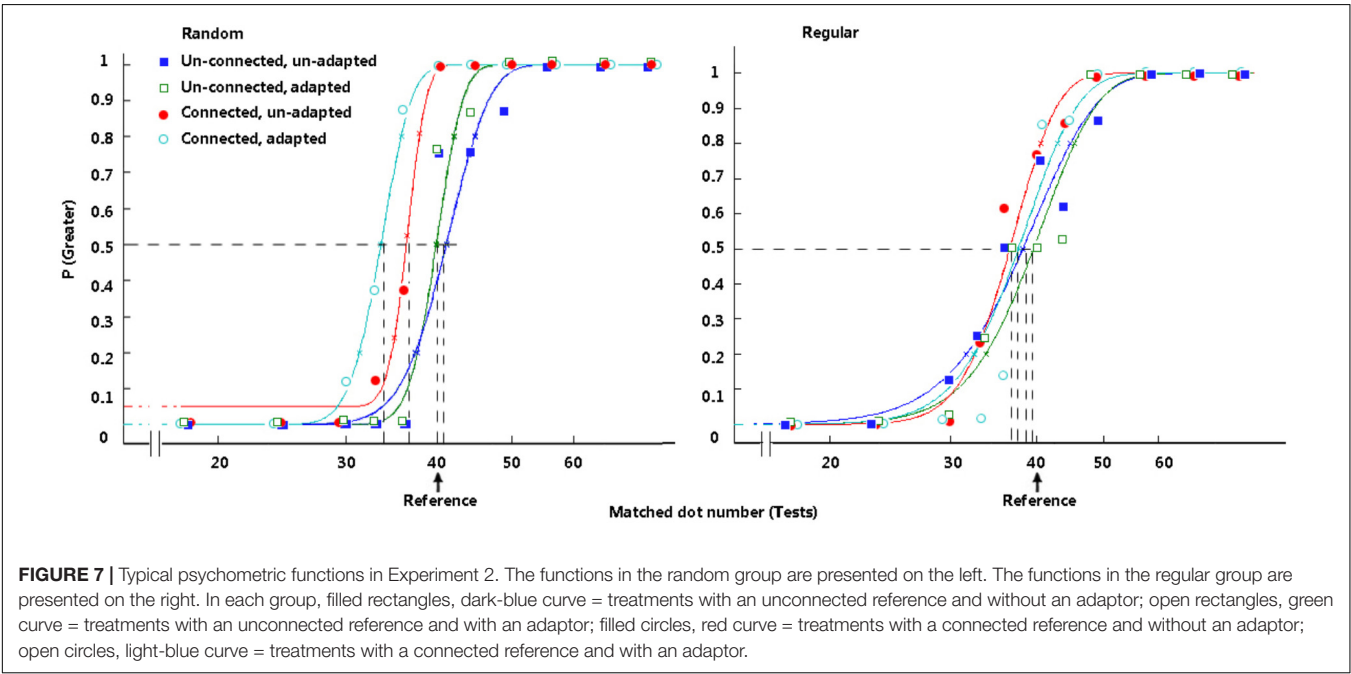


FIGURE 7 | Typical psychometric functions in Experiment 2. The functions in the random group are presented on the left. The functions in the regular group are presented on the right. In each group, filled rectangles, dark-blue curve = treatments with an unconnected reference and without an adaptor; open rectangles, green curve = treatments with an unconnected reference and with an adaptor; filled circles, red curve = treatments with a connected reference and without an adaptor; open circles, light-blue curve = treatments with a connected reference and with an adaptor.

TABLE 2 | The means and SDs for the PSEs in each treatment in Experiment 2.

Group	Random				Regular			
	Unconnected Unadapted	Unconnected Adapted	Connected Unadapted	Connected Adapted	Unconnected Unadapted	Unconnected Adapted	Connected Unadapted	Connected Adapted
PSE	40.92	40.00	36.28	34.36	39.47	39.62	37.62	38.53
SD	4.33	3.97	4.63	4.32	3.99	3.92	5.58	4.66

Unconnected/Connected: treatments in which the reference dots were unconnected/connected by lines. Unadapted/Adapted: treatments without/with the adaptation stage.

Compared with the results of the unconnected baseline, a significant difference was revealed when dots were connected by lines, $t(15) = 4.087$, $p = 0.001$, $d = 1.02$. Connection significantly decreased the perceived magnitude for random dots. Importantly, when the participants were comparing the number of dots connected by lines after adaptation, there was a further decrease for PSEs compared with the connected condition without adaptation, $t(15) = 2.585$, $p = 0.021$, $d = 0.65$. This difference indicates that adaptation affected perceived numerosity when the dots were connected in the reference, even though the dot number was equal in the adaptor and the reference. When the presented scene was shifted from the adaptor to the reference, connectedness decreased the perceived magnitude of the reference, and adaptation intensified the reduction in PSE. These results are in accord with previous research results (Fornaciai et al., 2016).

In the regular group, the situation seemed to be different. When dots were not connected, no significant PSE difference was found between circumstances with and without adaptation ($p = 0.829$). When dots were connected, the PSE difference between conditions with and without adaptation was not significant, either ($p = 0.312$).

A marginally significant decrease was found when the reference dots were connected (in the condition without adaptation) compared with the unconnected and unadapted baseline, $t(15) = 1.889$, $p = 0.078$, $d = 0.47$. Here, we provide a discussion of this marginal effect. When we compared the perceived numerosity of the treatments in which the lines were not controlled to be constant, the connectedness effect and/or the appearance of lines could be potential causes for the change in perceived numerosity. In our previous studies, in which the numbers rather than the distribution of lines were counterbalanced in the tests and the reference, the magnitude of the connectedness effect was directly related to the number of connected dot pairs in the random group (8 connected pairs, an 8-dot decrease in PSE), whereas the connection caused only a one- or two-dot decrease in PSEs in the three regular groups (Liu et al., 2017). The distinct magnitude of the decrease effect indicates that connection affects number perception in the random groups by changing numeral unit individuation; in contrast, number perception was affected because lines and connections caused a texture difference between the reference and the tests in the regular groups. In the current study, the decrease caused by lines is one or two dots in the regular group and approximately four dots in the random group. To some extent, the reduction still differs in the two groups, suggesting that the connection effect in the regular group acts differently from that in the random group. Nevertheless, the coding immunity of regular patterns regarding connectedness is mainly supported by that there was no significant difference between the connected treatments with and without adaptation.

DISCUSSION

The independence of numerosity processing from the processes associated with texture features, such as element size, orientation,

and texture, has been confirmed repeatedly by previous studies (Burr and Ross, 2008a,b; Liu et al., 2012, 2013, 2017). This independence demonstrates the involvement of abstraction processing in numerosity coding. In the current study, numerosity adaptation was shown to be independent of the change in element size in Experiment 1. This result is in accord with those of previous studies (Burr and Ross, 2008a; Burr, 2013).

In contrast, Experiment 1 demonstrated that the change in the element size relationship between adaptors and tests could affect numerosity adaptation for regular patterns. Adapting to large patterns and being tested with small patterns (the large-small condition) caused stronger aftereffects than adapting to small patterns and being tested with large patterns (the small-large condition). It is suggested that open space, which refers to the space that is not occupied by elements in a scene, is relevant to numeral comparison. The participants might have referred to non-numerical cues such as open space when they were asked to compare the numerosity of two sets of dots (Sophian, 2007). In Experiment 1, with an equal dot number, open space was inversely proportional to dot size. When the presented scene was shifted from the adaptor to the tests, the open space increased more dramatically under the large-small condition, in which the dot number decreased from 68 to less than 68 (in most cases) and the dot size transferred from 14×14 to 6×6 (pixel), than under the small-large condition, in which the dot number changed equally but the dot size changed inversely. Comparably, the adaptation aftereffect was revealed to be stronger under the large-small condition. We suggest that, for regular patterns, numerosity adaptation occurs via the adaptation of open space or open distance between adjacent dots. For regular patterns, numerosity estimation may use distance estimation as a reference. Abstraction seems to be inhibited, and texture specificity has been revealed repeatedly (orientation specificity, Liu et al., 2017; size specificity, the current study) in the numerosity coding of regular patterns.

Our previous studies suggested that regularity inhibited generic numerosity processing by inhibiting high-level processing, such as individuation. Experiment 2 provides new evidence for this suggestion. For random patterns, a reduction in magnitude perception was found when dots were connected, and a further reduction was revealed when the participants were asked to perceive the magnitude of the connected reference after adaptation to an adaptor whose dots were equal in number to those of the reference and were not connected. These results, which suggest that numerosity coding and adaptation directly affect perceptual mechanisms sensitive to number, are comparable to those of previous studies (Liu et al., 2012; Fornaciai et al., 2016). For regular patterns, however, the connectedness effect was absent in numerosity adaptation. This absence suggests an inhibition of individuation, which should be located in a higher step of numerosity processing and should be based on the activity of a set of complex neurons (Liu et al., 2017). Compared with the paradigm used in our 2017 study,

the adaptation paradigm in the current study provides improved evidence for the absence of a connectedness effect in the coding of regular patterns. Because the appearance of lines was kept constant in the treatments with and without adaptation, texture differences did not disrupt the comparison of perceived numerosity between these treatments.

Generic numerosity processing is likely to involve the activity of abstraction and individuation. When numerosity processing goes from a low to a high level, the primary coding of visual features could be discarded to form an abstract representation of the numerical units (Stoianov and Zorzi, 2011). Additionally, the magnitude estimation is likely based on the distinct number of items that have been individuated (Gallistel and Gelman, 2000). When high-level processing is inhibited by the visual properties of texture, numerosity processing may be indistinguishable from texture processing (Anobile et al., 2014; Liu et al., 2017). It is possible that regular distribution could cause a general inhibition for high-level processing in numerosity coding, including individuation and abstraction. The inhibition must function automatically rather than strategically, as no strategy was encouraged when the participants were asked to passively watch the screen, and they were informed that the adaptors were irrelevant to the tasks in our studies.

There might be a good reason for the inhibition of high-level numerosity processing in regularly distributed patterns. In natural scenes, it is more efficient to inhibit unnecessary (high-level) processing that achieves generic numerosity cognition when we observe regular patterns because it is more likely that we can obtain useful information by classifying “what” than by estimating “how many” (Liu et al., 2017).

More evidence suggesting that numerosity processing and texture processing share a common origin and arrive at distinct destinations could be gathered, for example, by comparing the event-related potential (ERP) component of numerosity and texture coding. Regardless, there will not necessarily be any contradiction in showing that various statistics of the image affect the approximation of numerosity. It is the distinguishable processing pertaining exclusively to numerosity coding, such as abstraction, individuation, unit representation, and spatially associated representation (Fischer, 2003; Liu et al., 2015), that determines the independent mechanism of numerosity cognition.

The inhibition caused by regular patterns, which was revealed repeatedly in our current and previous studies (Liu et al., 2017),

suggests an important role of random distribution in generic numerosity processing. Recently, a handful of studies have investigated common factors underlying approximate number system (ANS) acuity and mathematical achievement (Halberda et al., 2008; Chen and Li, 2014). The accurate measure of ANS acuity is important for this line of investigation. The current study provides additional suggestions on the design of tasks that measure the acuity of ANS. To measure the ANS acuity in an accurate manner, it is necessary to adopt a random dot pattern, as regularity in pattern distribution would inhibit generic numerosity coding. Similarly, it is also necessary to adopt a pattern with moderate density, as numerosity coding could also be inhibited by a cloudy-like effect (Anobile et al., 2014).

CONCLUSION

Dot size has a distinct effect on numerosity adaptation with random and regular distributed patterns. For random patterns, the change in stimulus size has no effect on adaptation. For regular patterns, adapting to large patterns and being tested with small patterns causes stronger aftereffects than adapting to small patterns and being tested with large patterns. The connectedness effect is different in the adaptation of random and regular patterns. For random patterns, references were perceived to be less numerous when the dots were connected via lines than when they were not connected, and there was a further underestimation of the connected references when the participants adapted to unconnected patterns with the same number of dots. This connectedness effect was absent in the numerosity estimation and the adaptation of regular patterns.

AUTHOR CONTRIBUTIONS

WL wrote the article and designed the experiments. YZ and ZZ edited the manuscript. MW collected the data.

FUNDING

This study was supported by funding from (1) the National Natural Science Foundation of China (Grant Nos. 31500879, 31371039, and 31500884) and (2) the Yunnan Science and Technology Planning Project (Grant No. 2017FB046).

REFERENCES

- Anobile, G., Cicchini, G. M., and Burr, D. C. (2014). Separate mechanisms for perception of numerosity and density. *Psychol. Sci.* 25, 265–270. doi: 10.1177/0956797613501520
- Anobile, G., Turi, M., Cicchini, G. M., and Burr, D. C. (2015). Mechanisms for perception of numerosity or texture-density are governed by crowding-like effects. *J. Vis.* 15, 1–12. doi: 10.1167/15.5.4
- Burr, D. C. (2013). “A visual sense of number,” in *Proceedings of the 9th Asia-Pacific Conference on Vision*, Suzhou.
- Burr, D. C., and Ross, J. (2008a). A visual sense of number. *Curr. Biol.* 18, 425–428. doi: 10.1016/j.cub.2008.02.052
- Burr, D. C., and Ross, J. (2008b). Response: visual number. *Curr. Biol.* 18, 857–858. doi: 10.1016/j.cub.2008.07.052
- Chen, Q. X., and Li, J. G. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychol.* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A., and Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19552–19557. doi: 10.1073/pnas.1113195108
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N

- Dehaene, S., and Changeux, J.-P. (1993). Development of elementary numerical abilities: a neuronal model. *J. Cogn. Neurosci.* 5, 390–407. doi: 10.1162/jocn.1993.5.4.390
- Dehaene, S., Izard, V., Spelke, E., and Pica, P. (2008). Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigenous cultures. *Science* 320, 1217–1220. doi: 10.1126/science.1156540
- Durgin, F. H. (2001). Texture contrast aftereffects are monocular; texture density aftereffects are binocular. *Vis. Res.* 41, 2619–2630. doi: 10.1016/S0042-6989(01)00121-3
- Durgin, F. H. (2008). Texture density adaptation and visual number revisited. *Curr. Biol.* 18, R855–R856. doi: 10.1016/j.cub.2008.07.053
- Fischer, M. H. (2003). Spatial representations in number processing: evidence from a pointing task. *Vis. Cogn.* 10, 493–508. doi: 10.1080/1350628024400186
- Fornaciai, G. M., Cicchini, D. C., and Burr, D. (2016). Adaptation to number operates on perceived rather than physical numerosity. *Cognition* 151, 63–67. doi: 10.1016/j.cognition.2016.03.006
- Franconeri, S. L., Bemis, D. K., and Alvarez, G. A. (2009). Number estimation relies on a set segmented objects. *Cognition* 113, 1–13. doi: 10.1016/j.cognition.2009.07.002
- Gallistel, C. R., and Gelman, R. (2000). Non-verbal numerosity cognition: from reals to integers. *Trends Cogn. Sci.* 4, 59–65. doi: 10.1016/S1364-6613(99)01424-2
- Halberda, J., Mazocco, M., and Feigenson, L. (2008). Individual differences in nonverbal number acuity predict maths achievement. *Nature* 455, 665–668. doi: 10.1038/nature07246
- Harris, J. J., Schwarzkopf, D. S., Chen, S., Bahrami, B., and Rees, G. (2011). Contextual illusions reveal the limit of unconscious visual processing. *Psychol. Sci.* 22, 399–405. doi: 10.1177/0956797611399293
- He, L. X., Zhang, J., Zhou, T. G., and Chen, L. (2009). Connectedness affects dot numerosity judgment: implications for configural processing. *Psychon. Bull. Rev.* 16, 509–517. doi: 10.3758/PBR.16.3.509
- He, L. X., Zhou, K., Zhou, T. G., He, S., and Chen, L. (2015). Topology-defined units in numerosity perception. *Proc. Natl. Acad. Sci. U.S.A.* 112, E5647–E5655. doi: 10.1073/pnas.1512408112
- Liu, W., Zhang, Z. J., Li, B. C., Zhao, Y. J., and Tang, Y. (2015). Dose numerosity adaptation activate MNLs? *Atten. Percept. Psychophys.* 77, 1358–1370. doi: 10.3758/s13414-015-0863-z
- Liu, W., Zhang, Z. J., and Zhao, Y. J. (2012). Numerosity adaptation effect on the basis of perceived numerosity. *Acta Psychol. Sin.* 44, 1297–1308. doi: 10.3724/SP.J.1041.2012.01297
- Liu, W., Zhang, Z. J., Zhao, Y. J., Li, B. C., and Wang, M. (2017). Distinct mechanisms in numerosity processing of random and regular dots. *Acta Psychol.* 174, 17–30. doi: 10.1016/j.actpsy.2017.01.006
- Liu, W., Zhang, Z. J., Zhao, Y. J., Liu, Z. F., and Li, B. C. (2013). Effects of awareness on numerosity adaptation. *PLoS One* 8:e77556. doi: 10.1371/journal.pone.0077556
- Milne, J. L., Chapman, C. S., Gallivan, J. P., Wood, D. K., Culham, J. C., and Goodale, M. A. (2013). Connecting the dots: object connectedness deceives perception but not movement planning. *Psychol. Sci.* 24, 1456–1465. doi: 10.1177/0956797612473485
- Raphael, S., Dillenburger, B., and Morgan, M. (2013). Computation of relative numerosity of circular dot textures. *J. Vis.* 13:17. doi: 10.1167/13.2.17
- Raphael, S., and Morgan, M. J. (2015). The computation of relative numerosity, size and density. *Vis. Res.* 124, 15–23. doi: 10.1016/j.visres.2014.12.022
- Sophian, C. (2007). “Measuring spatial factors in comparative judgments about large numerosities,” in *Foundations of Augmented Cognition: Third International Conference*, eds D. Schmorrow and L. Reeves (Secaucus, NJ: Springer), 157–165. doi: 10.1007/978-3-540-73216-7_18
- Stoianov, I., and Zorzi, M. (2011). Emergence of a ‘visual number sense’ in hierarchical generative models. *Nat. Neurosci.* 15, 194–196. doi: 10.1038/nn.2996
- Sweeny, T. D., Grabowecy, M., and Suzuki, S. (2011). Awareness becomes necessary between adaptive pattern coding of open and closed curvatures. *Psychol. Sci.* 22, 943–950. doi: 10.1177/0956797611413292
- Wichmann, F. A., and Hill, N. J. (2001). The psychometric function. I. Fitting, sampling, and goodness of fit. *Percept. Psychophys.* 63, 1293–1313. doi: 10.3758/BF03194544
- Zhang, Z. J., Liu, W., Zhao, Y. J., Zhang, J. S., and Wu, B. X. (2014). Cortical remapping features of numerosity adaptation aftereffects. *Acta Psychol. Sin.* 46, 5–16. doi: 10.3724/SP.J.1041.2014.00005

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Liu, Zhao, Wang and Zhang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



On the Difference Between Numerosity Processing and Number Processing

Anne H. van Hoogmoed^{1,2*} and Evelyn H. Kroesbergen^{1,3}

¹ Department of Pedagogical and Educational Sciences, Utrecht University, Utrecht, Netherlands, ² Department of Special Needs Education and Youth Care, University of Groningen, Groningen, Netherlands, ³ Behavioural Science Institute, Radboud University, Nijmegen, Netherlands

The ANS theory on the processing of non-symbolic numerosities and the ANS mapping account on the processing of symbolic numbers have been the most popular theories on numerosity and number processing, respectively, in the last 20 years. Recently, both the ANS theory and the ANS mapping account have been questioned. In the current study, we examined two main assumptions of both the ANS theory and the ANS mapping account. ERPs were measured in 21 participants during four same-different match-to-sample tasks, involving non-symbolic stimuli, symbolic stimuli, or a combination of symbolic and non-symbolic stimuli (i.e., mapping tasks). We strictly controlled the visual features in the non-symbolic stimuli. Based on the ANS theory, one would expect an early distance effect for numerosity in the non-symbolic task. However, the results show no distance effect for numerosity. When analyzing the stimuli based on visual properties, an early distance effect for area subtended by the convex hull was found. This finding is in line with recent claims that the processing of non-symbolic stimuli may be dependent on the processing of visual properties instead of on numerosity (only). With regards to the processing of symbolic numbers, the ANS mapping account states that symbolic numbers are first mapped onto their non-symbolic representations before further processing, since the non-symbolic representation is at the basis of processing the symbolic number. If the non-symbolic format is the basic format of processing, one would expect that the processing of non-symbolic numerosities would not differ between purely non-symbolic tasks and mapping tasks, resulting in similar ERP waveforms for both tasks. Our results show that the processing of non-symbolic numerosities does differ between the tasks, indicating that processing of non-symbolic number is dependent on task format. This provides evidence against the ANS mapping account. Alternative theories for both the processing of non-symbolic numerosities and symbolic numbers are discussed.

Keywords: number processing, ERP, ANS mapping account, non-symbolic, quantity processing, visual properties

OPEN ACCESS

Edited by:

Marcus Lindskog,
Uppsala University, Sweden

Reviewed by:

Dawei Li,
Duke University, United States
Robert Reeve,
The University of Melbourne, Australia

*Correspondence:

Anne H. van Hoogmoed
a.h.van.hoogmoed@rug.nl

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 13 April 2018

Accepted: 17 August 2018

Published: 12 September 2018

Citation:

van Hoogmoed AH and
Kroesbergen EH (2018)
On the Difference Between
Numerosity Processing and Number
Processing. *Front. Psychol.* 9:1650.
doi: 10.3389/fpsyg.2018.01650

INTRODUCTION

A prominent view on number processing is that non-symbolic quantities are processed intuitively by the approximate number system (ANS; Dehaene, 1997). The numerosity of a set of objects is assumed to be approximated by this system. This ANS theory is confirmed in a number of studies in infants, showing sensitivity to the numerosity of a set of objects from 6 months

of age (Xu and Spelke, 2000; Xu et al., 2005). Based on these studies, the processing of numerosity is assumed to be innate and shared across species (Xu and Spelke, 2000; Xu et al., 2005; Izard et al., 2009). Whereas the ANS theory concerns processing of the numerosity of sets of objects, an extension of the theory, named the ANS mapping account, is concerned with the processing of symbolic numbers. The ANS mapping account states that symbolic number processing is dependent on the ANS. Symbolic numbers that are encountered, are assumed to be first converted into a non-symbolic numerosity before further processing (Dehaene, 1997). Recently, both the ANS theory and the ANS mapping account have been questioned (Cohen Kadosh and Walsh, 2009; Gebuis et al., 2016; Lourenco et al., 2016; Reynvoet and Sasanguie, 2016; Leibovich et al., 2017; Núñez, 2017). The current study had two goals. First, we aimed to examine whether the processing of non-symbolic numerosity does indeed rely on an intuitive approximation of the numerosity of a set of objects, which would confirm the ANS theory. Second, we examined whether the processing of symbolic numbers is indeed based on the ANS as assumed by the ANS mapping account.

The ANS Theory

The ANS theory has been the most influential account on numerosity processing for the last 20 years. It suggests that the numerosity of a set of objects is approximated by extracting the numerosity from this set of objects independently of the visual properties of the set. Based on a mental number line, numerosities can be compared to each other (Dehaene, 1997). The approximation means that a set of objects does not only activate the corresponding numerosity, but also numerosities that are nearby on the mental number line. As such a set of 15 objects does not only activate the quantity 15 on the mental number line, but also 14 and 16, and to a lesser degree, 13 and 17. This leads to overlapping neural representations of the numerosities 15 and 16, but not for example 15 and 30. The larger the numerosity to be estimated, the more neighboring numerosities are co-activated. This explains why it is harder to distinguish between 15 and 16 objects than between 15 and 30 objects, and harder to distinguish between 15 and 16 than between 5 and 6.

Evidence for the ANS theory is mainly based on the results of comparison tasks. In these tasks, two sets of dots are presented and participants have to decide which set contains the largest number of dots. Lower accuracy and longer reaction times are obtained when the ratio between two quantities is closer to 1. For example, it is more difficult to compare 6 vs. 8 dots (ratio 0.75) than to compare 4 vs. 8 dots (ratio 0.5), but also more difficult to compare 6 vs. 8 dots (ratio 0.75) than to compare 4 vs. 6 dots (ratio 0.66). This effect is called the ratio effect (Reynvoet and Sasanguie, 2016; Smets et al., 2016) and is thought to be due to the co-activation of numerosities that are close on the number line. The closer the numerosities are to each other, the more they co-activate the same numerosity, which makes it more difficult to decide which is the larger one, in turn resulting in lower accuracy and higher reaction times. This ratio effect is not limited to behavioral studies, but is also shown in ERP research, where the amplitudes of the ERP signal differ per

ratio between two numerosities. More specifically, ERP studies on non-symbolic processing have shown ratio-dependent ERP amplitudes in varying time windows between 120 and 490 ms (Temple and Posner, 1998; Libertus et al., 2007; Paulsen and Neville, 2008; Hyde and Spelke, 2009, 2012). These ratio effects may reflect numerosity processing based on the ANS. However, the effects may also be due to the processing of the visual properties of the non-symbolic stimuli (i.e., a set of dots) instead of the numerosity of the sets.

In real life, visual properties of a set of objects co-vary with the number of objects in the set. For example, if you compare 5 fish to 10 fish, than the larger number of fish also occupies more of the visual scene, both in total surface of the fish as well as the area they occupy. Thus, in determining which group contains most fish, one could use both the visual properties (such as surface or area) as well as numerosity. The same holds for arrays of dots (or other non-symbolic stimuli). As such, it is difficult to distinguish the processing of visual input from the processing of numerosity. This problem has been acknowledged within the field for many years already (Mix et al., 2002). Different methods have been developed to control for visual input to be able to examine pure numerosity processing. Most ERP studies have used some sort of control for visual input when studying the processing of non-symbolic numerosities. An often-used method to control for effects of visual input has been described by Dehaene et al. (2005). Using this method, on half of the trials, the total surface of the dots or convex hull is equated, whereas the diameter of the dots and the distance between the dots varies. On the other half of the trials, diameter or distance between dots is equated, and total surface or convex hull varies. Studies using this type of control for visual input still show early ERP effects for small quantities (Libertus et al., 2007; Hyde and Spelke, 2012), which may suggest that numerosity processing is indeed automatic. However, these results may be due to the impossibility to strictly control for visual parameters when using small quantities. When using larger quantities, the early N1 effects disappeared, but distance effects were still found in the P2p time window, suggesting that numerosity is processed in a ratio-dependent manner in the latter time window (Libertus et al., 2007; Hyde and Spelke, 2012).

Gebuis and Reynvoet (2011) suggested that the control for visual input developed by Dehaene et al. (2005) may not be sufficient. Participants could not rely on a single visual property to compare numerosities, but could still use total surface or convex hull in half of the trials, and diameter or distance between the dots in the other half of the trials. Therefore, Gebuis and Reynvoet (2011) developed a more advanced method to control for visual properties in which all properties are varied simultaneously and visual properties only explain a very small portion of the variance in numerical distance (Gebuis and Reynvoet, 2011). When comparing this method with a method similar to the one developed by Dehaene et al. (2005), diverging results were found (Gebuis and Reynvoet, 2012). When using the method of Dehaene et al. (2005) N1 and P2 effects were found. When controlling for visual input with the method developed by Gebuis and Reynvoet (2011), no N1 and P2 effects were found, suggesting that the N1 and P2 effects found in the first experiment are explained by visual cues. Also other studies using this more

stringent method of Gebuis and Reynvoet (2011) found distance effects only in later ERP components starting around 600 ms (Soltész and Szűcs, 2014), or no ERP components related to distance at all (Gebuis and Reynvoet, 2013). This suggests that the processing of non-symbolic stimuli is not based on the extraction of approximate numerosity, but instead relies on the processing of visual features.

Indeed the ANS has recently been questioned based on the abovementioned results (Gebuis et al., 2016; Leibovich et al., 2017; Núñez, 2017), and alternatives have been proposed. Gebuis et al. (2016) propose a sensory integration theory, in which visual properties are not removed in order to compare numerosity, but are the basis of this comparison (see also Gevers et al., 2016). Different sensory cues are integrated to compare numerosities. Related to this theory, Leibovich et al. (2017) propose a sense for magnitude theory instead of a sense for number. This theory states that magnitude processing and not number processing is automatic and innate. They claim that the development of numerosity processing is based on this sense for magnitude as children discover the relation between numerosity and magnitude. However, several comments on this paper counter this idea by arguing that a sense of numerosity is innate and automatically extracted, as also posed by the ANS theory (Content et al., 2017; de Hevia et al., 2017; Libertus et al., 2017; Nieder, 2017; Park et al., 2017; Savelkoul and Cordes, 2017; Stoianov and Zorzi, 2017).

In the current study, we aimed to give further insight into the processing of non-symbolic numerosities. Therefore, we examined the timing of ratio-related distance effects in the ERP while using larger quantities and stringent control over visual properties by using the method of Gebuis and Reynvoet (2011). Based on the ANS theory, one would expect early ratio-related distance effects in the ERP, suggesting processing of numerosity independent of visual properties. However, an absence of early effects for numerosity in combination with longer lasting effects based on visual properties, would suggest that visual properties of stimuli are not removed to approximate numerosity, but visual properties do play a role in determining numerosity. An absence of the ratio-related distance effect would support the previous findings discussed above (Gebuis and Reynvoet, 2012, 2013; Soltész and Szűcs, 2014). However, these studies examined passive viewing of dot patterns (Gebuis and Reynvoet, 2012, 2013; Soltész and Szűcs, 2014), in which the attention of the participants was not directed toward the numerosity of the set. Only in the second experiment in the study of Gebuis and Reynvoet (2013), participants were instructed to attend to the numerosity by including attention trials on which the participant needed to estimate the numerosity of the current stimulus. However, manipulation of the distance or ratio between two stimuli, as more generally used in ERP and behavioral research on numerosity processing (Moyer and Landauer, 1967; Temple and Posner, 1998; Libertus et al., 2007; Paulsen and Neville, 2008; Hyde and Spelke, 2009) is lacking.

The ANS Mapping Account

The ANS is not only the most prominent theory on non-symbolic number processing, but also the basis for the most

common model for the processing of symbolic numbers. This model on symbolic number processing based on the ANS is referred to as the ANS mapping account. The core of the ANS mapping account is that adults intuitively map symbolic numbers onto the corresponding non-symbolic numerosity before further processing (Dehaene, 1997). As such, a comparison task with symbolic stimuli is solved in a manner similar to a non-symbolic comparison task after mapping the symbolic number onto the non-symbolic numerosity.

The ANS mapping account is supported by symbolic comparison tasks that show effects similar to the ratio effect found for non-symbolic stimuli. More specifically, behavioral performance on symbolic comparison tasks reflects distance and size effects (Dehaene et al., 1990; Verguts and Van Opstal, 2005; Holloway and Ansari, 2008; Sasanguie et al., 2012, 2013). The distance effect entails better performance when two quantities are further apart from each other, whereas the size effect entails better performance for small numerosities as compared to large numerosities when the distance between them is equal (i.e., 3 vs. 4 is easier to compare than 7 vs. 8). Together, the distance and size effects are similar to the ratio effect found in non-symbolic comparison tasks (Holloway and Ansari, 2008; Halberda et al., 2012; Sasanguie et al., 2012, 2013), which is thought to support the ANS mapping account (see Reynvoet and Sasanguie, 2016 for a review). ERP studies have shown that the timing of these effects is also similar to the ratio-effects found in non-symbolic processing (Dehaene, 1996; Temple and Posner, 1998; Libertus et al., 2007). Together, these results suggest that the processing of symbolic number relies on the processing of non-symbolic numerosity.

However, the underlying assumption that distance effects found in behavioral and ERP research reflect overlapping neural representations has been questioned. Research has shown that the distance effect found in comparison tasks, hence called the comparison distance effect (CDE), does not necessarily originate from the larger overlap in neural representation in two numerically close numbers, but may be caused by more general decision processes (Van Opstal et al., 2008). Comparison tasks with letters and digits were compared to each other. Participants had to indicate whether a digit between 1 and 9 was smaller or larger than 5, and whether a letter between J and R came either before or after the letter N in the alphabet. A CDE was found for both letters and digits, even though letters are not assumed to have overlapping neuronal representations with neighboring letters, suggesting that the distance effects found in comparison tasks do not necessarily support the ANS mapping account.

In the same paper, Van Opstal et al. (2008) re-analyzed the data from the comparison task based on the distance between the previous digit or letter (the prime) and the current number or letter (the target). They showed that reaction times were shorter when the digit in the previous trial was close to the digit presented in the current trial (4 preceded by 3) than when the digit in the previous trial was further away from the one presented in the current trial (4 preceded by 1). This faster reaction is assumed to be due to the fact that the quantity was already partly activated, and thus primed, during processing of the previous digit, and hence named the prime distance effect (PDE). This effect was

found to be specific for digits, and not present for letters. In a follow-up study, van Opstal and Verguts (2011) found that the PDE was not limited to the specific task described above, but could also be found in a same-different match-to-sample task. In this task, participants were presented with two symbolic numbers (a digit and a number word) consecutively and had to respond to indicate whether these stimuli depict the same or a different quantity. For stimuli that differed from each other, the distance between the prime (number that is presented first) and the target (number that is presented second) was manipulated. Reaction times to the “different” targets were faster when the numbers were further apart from each other (e.g., 2 vs. 8) than when the numbers were close to each other (e.g., 7 vs. 8). This was interpreted as an effect of more co-activation due to overlapping neural representations in the latter case. However, this study did not examine whether these distance effects for symbolic numbers were related to distance effects found for non-symbolic stimuli. Behavioral evidence shows low correlations between the distance effects in symbolic and non-symbolic tasks, questioning whether these tasks are solved based on similar processing in both tasks (Holloway and Ansari, 2009). Also, recent research shows that although the ANS model can describe behavioral results in non-symbolic tasks relatively well, it has difficulty in describing behavioral results in symbolic comparison tasks, again indicating that symbolic numbers are not processed by the ANS (Krajcsi et al., 2018).

To directly investigate similarities between symbolic and non-symbolic processing, mapping tasks in which symbolic and non-symbolic quantities need to be compared to each other should be used. Based on the ANS mapping account that symbolic processing is rooted in non-symbolic numerosity processing, one would expect that results in purely symbolic tasks, purely non-symbolic tasks, and tasks in which symbolic and non-symbolic numbers need to be combined are similar. More specifically, one would expect that the processing of non-symbolic numerosities would not be affected by the format of the stimulus it needs to be compared to. As such, based on the ANS mapping account, one would not expect differences between the primes in the purely non-symbolic task and the mapping task with non-symbolic primes and symbolic targets. Similarly, one would not expect differences between the processing of non-symbolic targets in the purely non-symbolic task and the mapping task with symbolic primes and non-symbolic targets. Behavioral evidence from mapping tasks shows that performance on mapping tasks is worse than performance on a purely non-symbolic comparison task (Lyons et al., 2012), suggesting that the mapping of symbolic numbers onto non-symbolic numerosities is not an intuitive process. Another study showed that tasks involving non-symbolic stimuli elicit a ratio effect, both completely non-symbolic tasks as well as when mapping tasks. However, purely symbolic tasks did not show a ratio effect. This suggests that the non-symbolic numerosity in the ANS may not be activated when comparing two symbolic numbers (Sasanguie et al., 2017).

These data question the validity of the ANS mapping account in two ways. First of all, they question whether symbolic numbers are mapped onto non-symbolic numerosities when this is not necessary for the task at hand. Second, they question

whether the possible mapping occurs intuitively. Therefore, in the current study, we measured ERPs in same-different match-to-sample tasks with symbolic stimuli, non-symbolic stimuli, or a combination between symbolic and non-symbolic stimuli to examine whether symbolic numbers are indeed mapped onto non-symbolic numerosities, and if so, whether this mapping is an automatic process. The ANS mapping account is examined in two ways. First, based on the ANS mapping account, one would expect similar distance effects in symbolic and mapping tasks as in the non-symbolic task if symbolic numbers are indeed mapped onto the ANS. Second, one would expect that the non-symbolic stimuli are processed similarly resulting in similar ERPs, regardless of whether they need to be compared to symbolic stimuli or non-symbolic stimuli, since the ANS is the core system, which is at the basis of numerical processing. Stated otherwise, a difference in the ERPs for non-symbolic stimuli depending on the task suggests that this does not lie at the basis of numerical processing. This would provide evidence against the ANS mapping account.

MATERIALS AND METHODS

Participants

Twenty-three adults, mainly undergraduate students, participated in the study. Two were excluded due to noisy EEG data (see below). The final sample consisted of four males and 17 females, with a mean age of 23 years and 10 months (SD 3 years, 3 months). Of the participants, 19 were right handed, and 2 were left handed. All participants had normal or corrected-to-normal vision. All participants gave written informed consent in accordance with the Declaration of Helsinki.

Procedure

Participants were seated in an electrically shielded room. They were informed that the study would assess numerical skills and consisted of four comparison tasks. Upon successful application of the EEG, the task instruction of the first task was presented on the screen. Participants were told that there would be a break after each task. During these breaks the researcher would come in to ask how they were doing and to answer any questions. The tasks were presented in a fixed order with the non-symbolic task first, then the non-symbolic/symbolic task, then the symbolic/non-symbolic task, and finally the symbolic task. The order of the tasks was fixed such that participants did not know which numbers were presented in the non-symbolic format and were not able to calculate the ratios based on the purely symbolic task. After the four tasks, the EEG cap was removed from the participant and they were financially compensated for participation with 10 Euros. All tasks including application and removal of the EEG-cap lasted about 75 min.

Tasks

Non-symbolic (Ns-Ns)

In the non-symbolic task, trials consisted of a prime picture with a dot pattern and a target picture with a dot pattern, see **Figure 1**. The dot patterns were generated in MATLAB with the

script described in Gebuis and Reynvoet (2011). Using this script, the relation between the number distance and visual properties was controlled, as well as the congruency in area subtended, density, total surface of the dots, average diameter, and total circumference. Moreover, visual properties of the stimuli are documented, which gives the opportunity to divide data based on visual properties as well (Gebuis and Reynvoet, 2011). The number of dots for the primes ranged between 20 and 40, with both smaller and larger targets at ratio 0.5, 0.6, and 0.7. As such, all numbers ranged between 10 and 80, and thus far out of the subitizing range. A trials started with the presentation of a prime for 750 ms, then a blank screen jittered between 400 and 600 ms, and a target presented for 750 ms. The inter trial interval was jittered between 1,000 and 1,500 ms. Thirty trials were presented for each distance \times size (target larger vs. target smaller than prime). In 10 percent of the trials (20 trials), the numerosity in the prime and the target were the same, resulting in a total of 200 trials¹. Participants were instructed to passively watch the stimuli and only respond by pressing the space bar if they thought the prime and target stimuli displayed the same quantity.

Non-symbolic – Symbolic (Ns-S)

The Ns-S task was identical to the Ns-Ns task with the exception that the targets were presented as digits instead of dot patterns.

Symbolic – Non-symbolic (S-Ns)

The S-Ns task was identical to the Ns-Ns task with the exception that the primes were presented as digits instead of dot patterns.

Symbolic (S-S)

The S-S task differed slightly from the Ns-Ns task. Both the prime and the target were presented as digits. Moreover, the stimuli were presented for 500 ms instead of 750 ms, since the task was very simple.

Analyses

Behavioral

Participants had to respond only to trials in which they thought the prime and target matched each other. As such, a non-response to the trials in which the prime and target did not match each other is taken as a correct response. Behavioral data were analyzed in SPSS, version 23. Proportions correct were analyzed per task in a Ratio (0.5, 0.6, and 0.7) \times Size (target larger vs. target smaller) repeated measures ANOVA. Polynomial contrasts were included to test whether performance increased linearly with decreasing ratio.

ERP

Recording and Preprocessing

Data were recorded with a 32 electrode active cap (Biosemi, Amsterdam, Netherlands) with a sampling rate of 2048 Hz. The electrode offset was kept below 50 μ V. Data were recorded without reference. After recording, data were imported

into MATLAB 2015a (The MathWorks Inc., Natick, MA, United States) and analyzed using the Fieldtrip toolbox (Oostenveld et al., 2011).

Data were downsampled to 512 Hz, rereferenced to the linked mastoids, and low-pass filtered at 40 Hz. ICA was used to identify and delete eye blinks and horizontal eye movements. After that, data were manually inspected for bad channels. Bad channels were removed and replaced with a weighted sum of the surrounding channels. Removed channels were never adjacent to each other. Data (primes and targets) were segmented from 200 ms before to 750 ms after stimulus onset and baseline corrected. After artifact rejection, the data were averaged per ratio per task for the targets and averaged per task for the primes. Data from target larger than prime and target smaller than prime were collapsed because of the limited number of trials included. Next to that, averages were generated for small, medium, and large diameter; small, medium, and large area; and small, medium, and large surface. The averages were created such that they contained the same number of trials as the averages per ratio.

Analyses

Single-subject averages were included in the analyses if at least 40 artifact free trials were included in the average for each condition. Since the time course of the differences between conditions was unknown, cluster based permutation tests were carried out. For the Ratio effects in the tasks, four separate permutation tests were carried out, one for each task. A linear effect of Ratio was expected. Therefore, the *t*-statistic of the slope of a multilevel linear estimation procedure with fixed slope and random intercept was used as input for the analyses. Similar cluster based analyses were performed for the physical parameters (mean) diameter, area (within the convex hull), and total surface (of the dots).

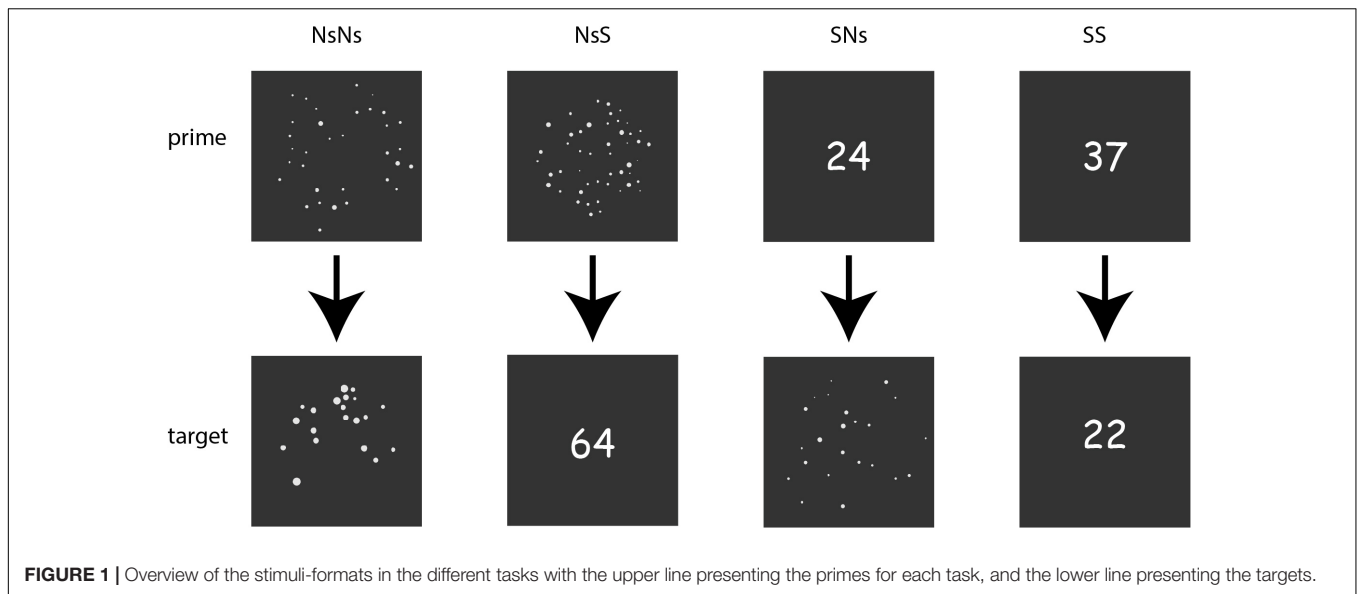
To test for differences in the processing of non-symbolic stimuli depending on task, two cluster based permutation tests were carried out, one to compare the processing of primes in the NsNs-task vs. the NsS-task, and one to compare the processing of the targets in the NsNs-task vs. the SNs-task. A dependent-samples *t*-test was used as input for the cluster based permutation test. Since cluster-based statistics (clusterstats) are calculated for positive and negative clusters separately, the *p*-values will be compared to $\alpha = 0.025$ (0.05/2) for all analyses.

RESULTS

Behavioral

Accuracy data for each task are presented in **Figure 2**. For the NsNs-task, the repeated measures ANOVA with the factors Ratio (0.5, 0.6, and 0.7) and Size (target smaller and target larger) revealed a main effect of Ratio, $F(2,40) = 53.03$, $p < 0.001$, but no significant effect of Size, $F(1,20) = 0.36$, $p = 0.554$, and no interaction between Ratio and Size, $F(2,40) = 1.89$, $p = 0.165$. The polynomial contrasts showed a linear trend for Ratio, $F(1,20) = 75.12$, $p < 0.001$, but no quadratic trend, $F(1,20) = 0.34$, $p = 0.569$. For the NsS-task the results were

¹ Stimulus generation failed for 18 out of 400 stimuli. As such, instead of 200 trials, 184 trials were presented in the Ns-Ns task, 187 trials in the Ns-S task, and 195 trials in the S-Ns task.



similar. A main effect of Ratio was found, $F(2,40) = 33.54$, $p < 0.001$, but no effect of Size, $F(1,20) = 1.86$, $p = 0.188$, and no interaction between Ratio and Size, $F(2,40) = 0.14$, $p = 0.803$. The polynomial contrasts indicated a linear trend as well as a quadratic trend, $F(1,20) = 38.71$, $p < 0.001$ and $F(1,20) = 4.92$, $p = 0.038$, respectively. This indicates that accuracy increases with the smaller ratio's, and the difference in accuracy is larger between 0.6 and 0.7 than between 0.5 and 0.6. For the SNs-task, a main effect of Ratio, $F(2,40) = 4.00$, $p = 0.049$, and a main effect of Size were found, $F(1,20) = 8.30$, $p = 0.009$. No interaction between Ratio and Size, $F(2,40) = 0.95$, $p = 0.361$ was present. The results show higher accuracy when the target was smaller than the prime as compared to when the target was larger than the prime. The polynomial contrasts indicated marginally significant linear and marginally significant quadratic trends, $F(1,20) = 4.03$, $p = 0.058$ and $F(1,20) = 3.76$, $p = 0.067$. In the SS-task, no significant main effects of Ratio and Size, $F(2,40) = 0.77$, $p = 0.470$ and $F(1,20) = 1.88$, $p = 0.186$, respectively, and no interaction between Ratio and Size were found, $F(2,40) = 0.59$, $p = 0.560$.

Ratio Effects Targets

ERPs depicting the ratio effects of the targets in the different tasks are shown in **Figure 3**. The results of the permutation test on the ratio effect in the Ns-Ns task shows no significant cluster for ratio, largest positive clusterstat = 1376.4, $p = 0.846$, and largest negative clusterstat = -4899.1, $p = 0.094$. For the NsS task, no significant cluster for ratio was found either, largest positive clusterstat = 2529.4, $p = 0.902$, largest negative clusterstat = -4985.4, $p = 0.246$. For the SNs task, no significant cluster for Ratio was found, largest positive clusterstat = 2459.7, $p = 0.816$, largest negative clusterstat = -2967.3, $p = 0.339$. For the SS task, results showed no significant clusters either, largest positive clusterstat = 2779.7, $p = 0.994$, largest negative clusterstat = -587.4, $p = 0.118$. These results reflect an absence of an effect for Ratio for all tasks.

Differences Between Tasks

Since no ratio effects were found, the timing of the ratio effects in the different tasks could not be compared. Hence, differences between tasks were only assessed based on the differences in the processing of the non-symbolic stimuli.

First, the processing of the primes in the NsNs-task and the NsS-task was compared. ERPs of the primes in these tasks are depicted in **Figure 4A**. The results of the permutation test on the primes in the NsNs and NsS task revealed a significant negative cluster, clusterstat = -3777.7, $p = 0.022$, but no significant positive cluster, largest clusterstat = 627.3, $p = 0.060$. The negative cluster reflects a fronto-central negativity between 125 and 400 ms, being relatively widespread between 125 and 175 ms, moving to mainly left-frontal between 275 and 400 ms (see **Figure 5**).

Second, the processing of targets in the NsNs-task was compared to the processing of the targets in the SNs-task. ERPs depicting the processing in both tasks are shown in **Figure 4B**. The permutation test on the difference between non-symbolic targets in the NsNs task and SNs task shows a significant positive cluster, clusterstat = 4243.8, $p = 0.012$ reflecting a right-frontal difference between 600 and 750 ms, and a significant negative cluster, clusterstat = -7613.6, $p = 0.002$ reflecting a widespread fronto-central negativity between 150 and 250 ms (see **Figure 6**).

Visual Properties of Non-symbolic Stimuli

The ERPs of the visual properties are displayed in **Figure 7**. With regards to the visual properties, the results of the permutation test on area showed a positive cluster, clusterstat = 24642, $p = 0.008$, but no significant negative cluster, largest clusterstat = -2583.8, $p = 0.571$. This cluster reflects a widespread positivity increasing with area covered between 200 and 750 ms in fronto-central to parietal regions (see **Figure 8**). The results of the permutation test on diameter showed no significant positive and negative cluster, with the largest clusters being, respectively, clusterstat = 17500,

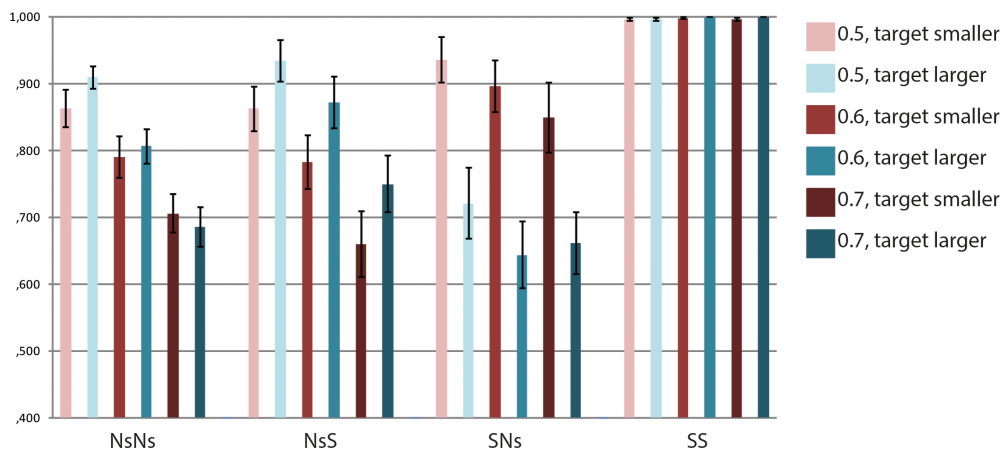


FIGURE 2 | Accuracy data per Ratio \times Size for each of the tasks.

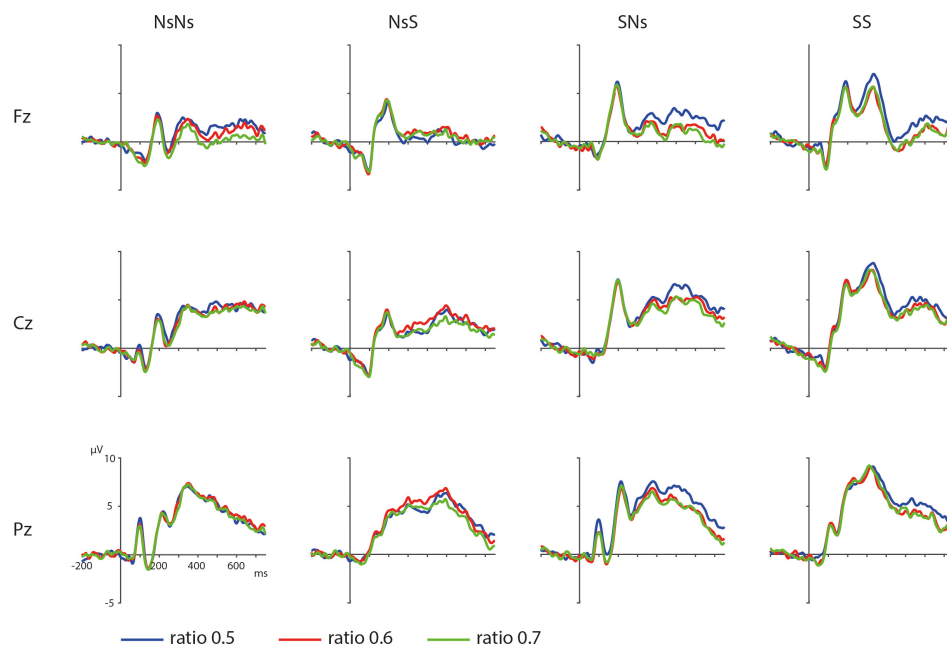


FIGURE 3 | Distance effects on frontal, central, and parietal electrodes in the different tasks with the blue waveform depicting targets with a ratio of 0.5, red depicting a ratio of 0.6, and green depicting a ratio of 0.7.

$p = 0.108$ and clusterstat = -2241.4 , $p = 0.465$. The results of the permutation test on surface show no significant positive cluster, largest clusterstat = 2298.4 , $p = 0.082$, and no significant positive cluster, largest clusterstat = -4005.7 , $p = 0.353$.

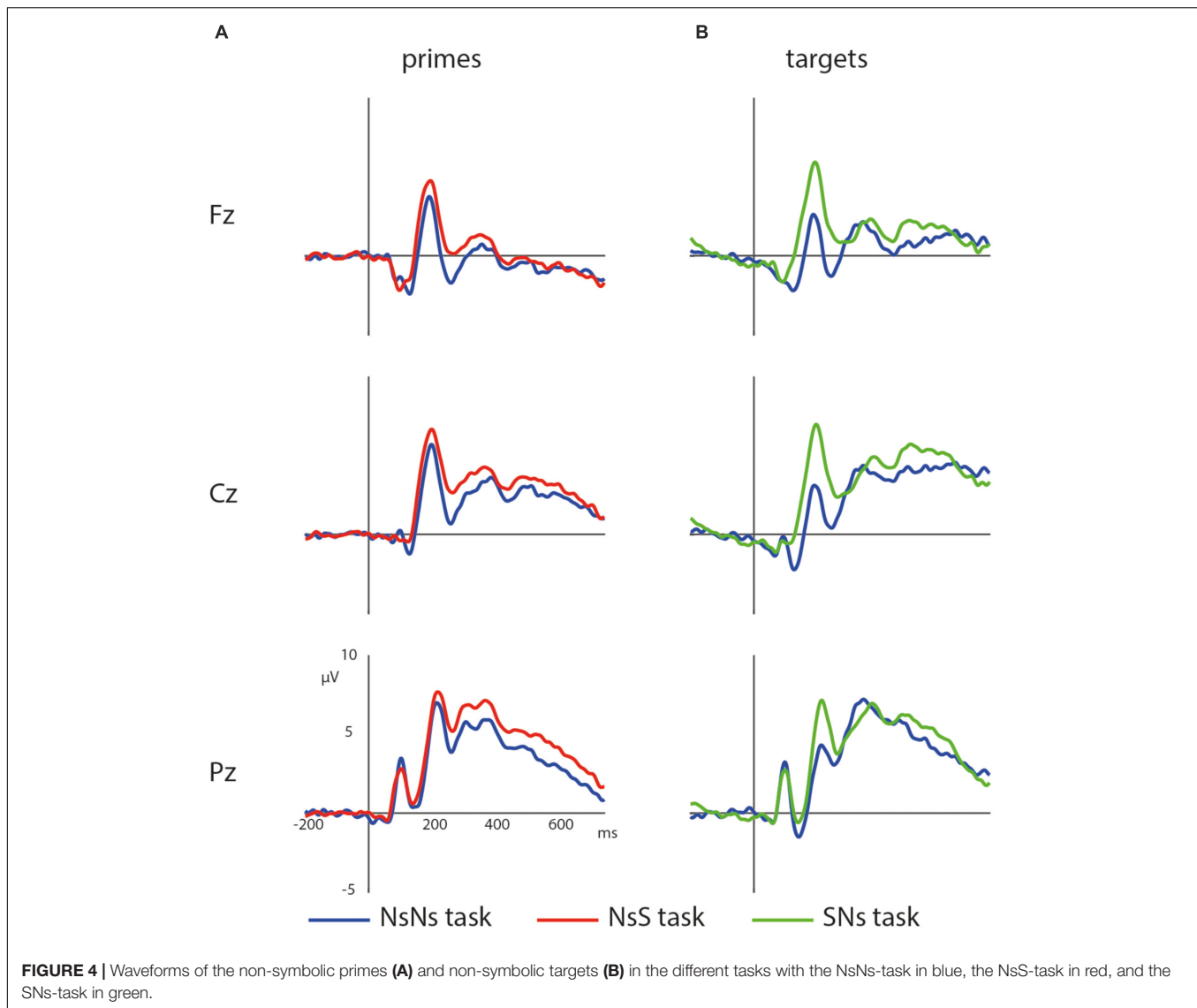
DISCUSSION

The ANS theory and ANS mapping account (Dehaene, 1997) have been the most prominent theories on number processing in the past decades. However, recently, the validity of the ANS theory and ANS mapping account have been questioned. The aim of the current study was twofold. First, we examined

whether non-symbolic numerosity is processed intuitively and independent of the processing of visual features as claimed by the ANS theory. Next, we examined whether symbolic numbers are mapped onto non-symbolic numerosities, as expected based on the ANS mapping account. ERPs were measured during a same-different match-to-sample task with non-symbolic numerosities, a task with symbolic numbers, and mapping tasks in which the prime was symbolic and the target non-symbolic or vice versa.

ANS Theory

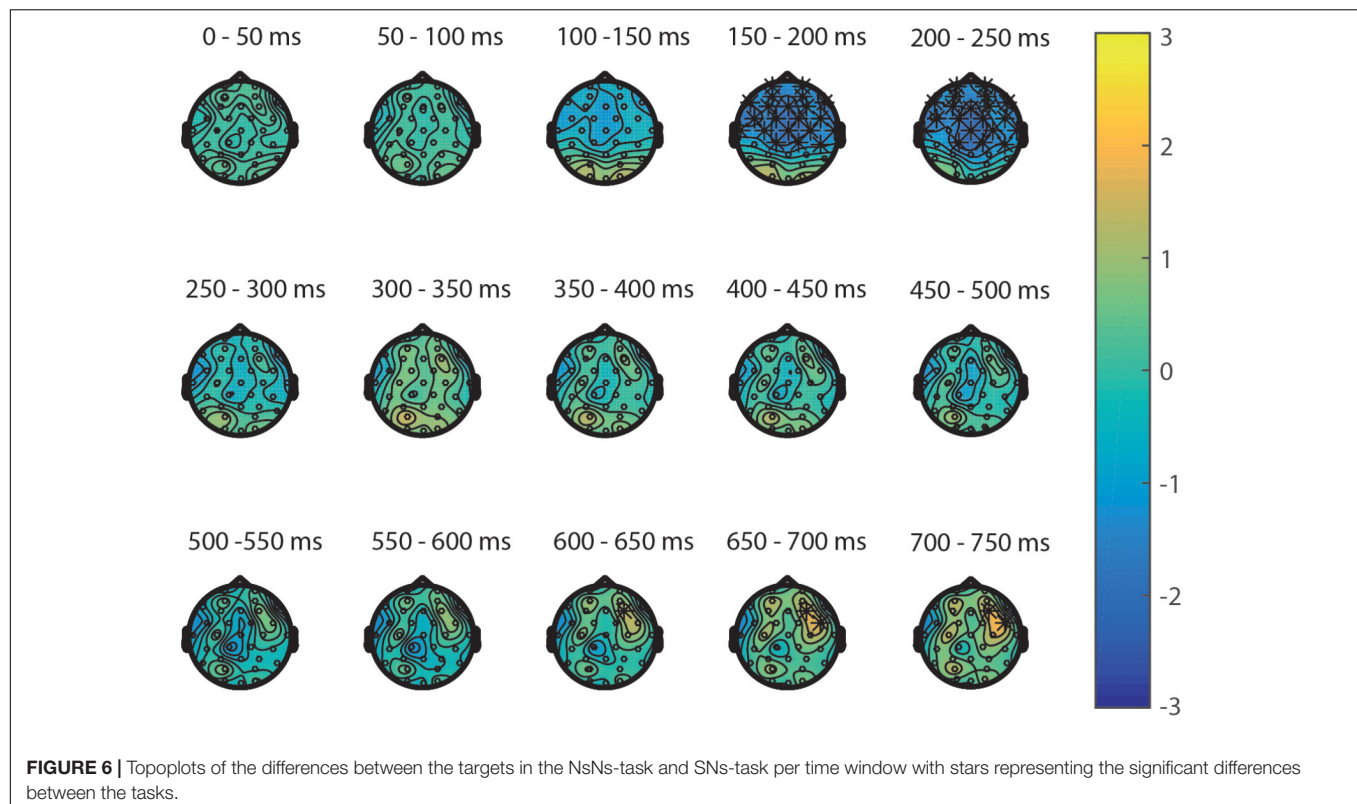
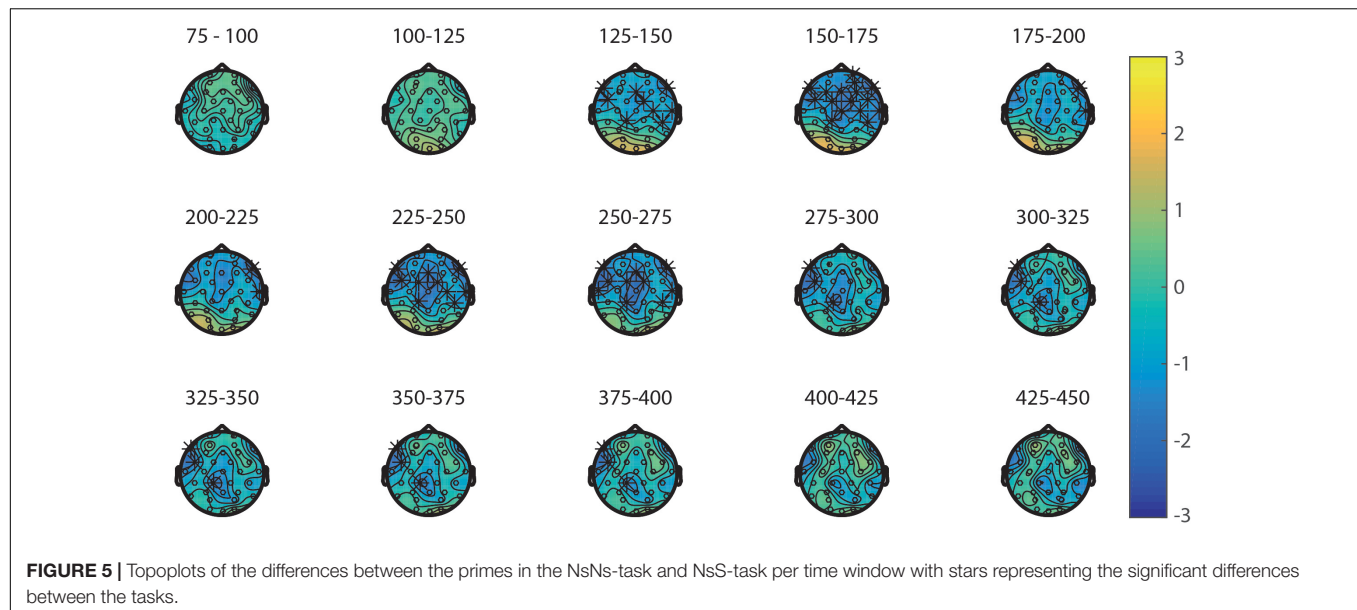
As support for the ANS theory, one would expect (early) distance effects in the completely non-symbolic task. Our results show



that despite the distance effect in the behavioral data, no ERP distance effects for numerosity were found, which means that the ratio between the numerosity of the prime and target was not visible in the ERP signal. This result is in line with previous research using strict control over visual properties (Gebuis and Reynvoet, 2012), and suggests that numerosity is not intuitively activated in non-symbolic stimuli. In contrast, the ERP results do show an early distance effect starting at 200 ms when stimuli are categorized based on the visual property area instead of numerosity, indicating that area is processed very quickly. This suggests that the area subtended by the convex hull around the dots is activated and processed. These results are in contrast with previous research in which processing of numerosity was claimed based on numerosity-related distance effects with non-symbolic stimuli (Temple and Posner, 1998; Libertus et al., 2007; Paulsen and Neville, 2008; Hyde and Spelke, 2009, 2012). In those studies, visual properties were not controlled for in a strict manner, resulting in the possibility to use visual properties to

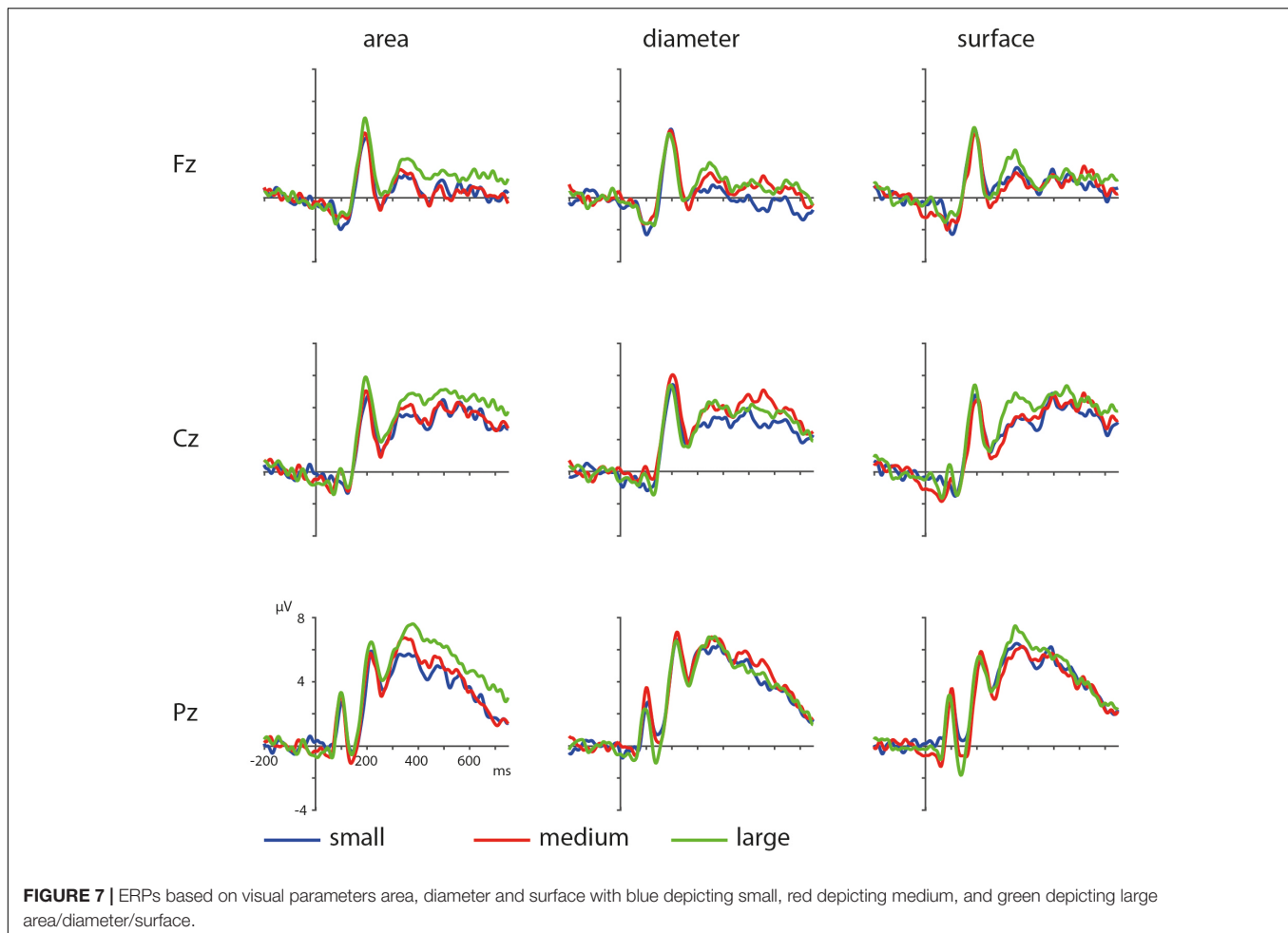
inform oneself about numerosity. In studies with proper control, Gebuis and Reynvoet (2012, 2013) also found effects for visual processing, but not for numerosity processing. This confirms that the early effects found in the abovementioned studies are likely due to insufficient control over visual properties, as suggested by Gebuis and Reynvoet.

An alternative explanation for the lack of a distance effect for numerosity is that the ANS theory does hold, but that this distance effect cannot be measured with ERP. Most models on the ANS theory suggest that individual objects go through a normalization phase in which sensory properties are removed before they enter the accumulator stage in which the information is transformed into numerosity (Dehaene and Changeux, 1993). Whereas a lack of a distance effect for numerosity in the ERP does not necessarily contradict to this idea, the presence of a long lasting distance effect for area, up until 750 ms, does. If the stimuli would go through a normalization phase, one would expect only effects of visual properties before this stage,



i.e., only very early in the ERP. Thus, our data support the claim that a normalization phase is unlikely (Gebuis et al., 2016). Taken together, our ERP results do not support the ANS theory. However, the behavioral distance effect suggests that approximate numerosity is established. Our ERP results suggest that this is achieved based on the processing of the visual properties. This is in line with previous research showing that visual properties are processed more automatically

as compared to numerosity (Gebuis and Reynvoet, 2013; Smets et al., 2015). As alternatives for the ANS theory, the sensory integration theory and sense of magnitude theory have been proposed (Gebuis et al., 2016; Leibovich et al., 2017). Our results with large and long-lasting distance effects for area and not numerosity, support these theories by showing that magnitude (in this case area) is processed more automatically than numerosity.

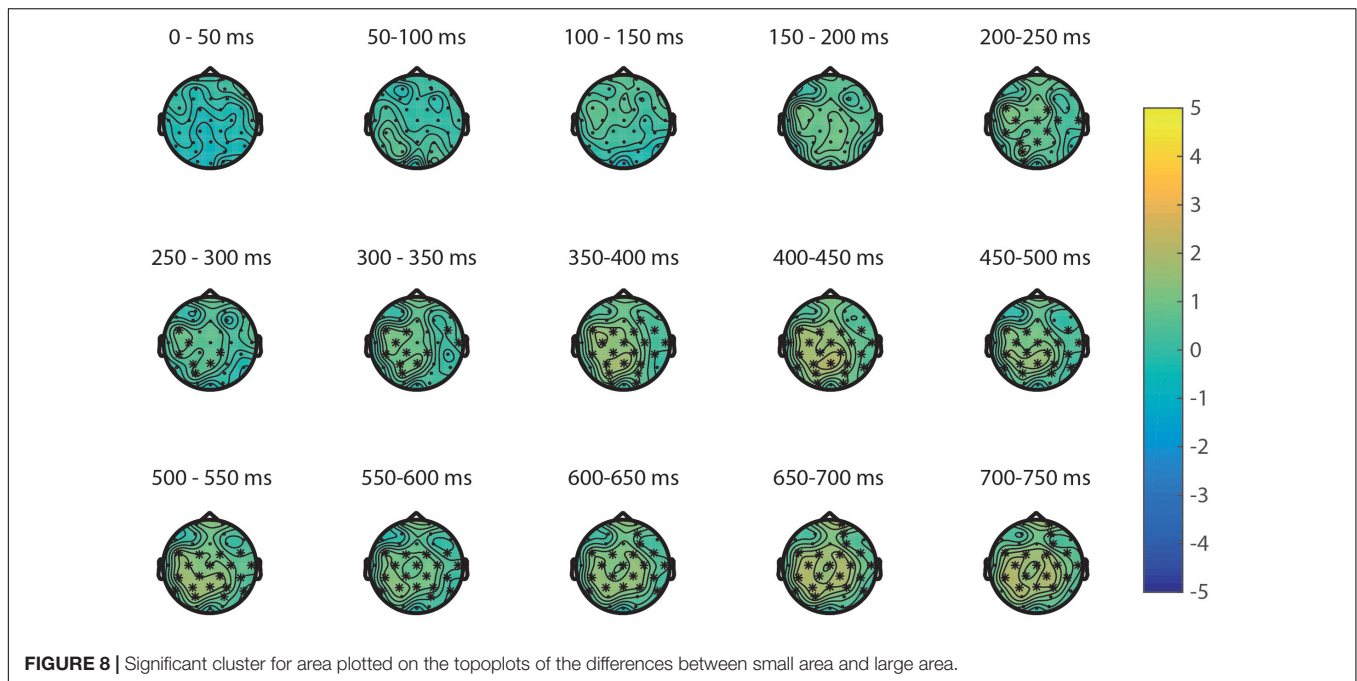


ANS Mapping Account

The second aim of our study was to examine the ANS mapping account. Whereas the results of the non-symbolic task question the existence of the ANS theory in its current form, mapping of symbolic stimuli onto their non-symbolic counterparts may still occur. The first line of evidence for the ANS mapping account would come from similar distance effects in the non-symbolic task and the symbolic and mapping tasks. The behavioral results shows similar distance effects in the non-symbolic and mixed tasks, but no distance effect in the purely symbolic task, which is in line with recent research (Sasanguie et al., 2017). This strengthens the claim that indeed in purely symbolic tasks, non-symbolic numerosity is not activated. The ERPs showed no distance effect in any of the tasks. Due to the lack of distance effects in the ERP, comparing these ERP distance effects between tasks is not possible.

The second line of evidence for the ANS mapping account would come from similar processing of non-symbolic stimuli regardless of task. Distance effects are no prerequisite to examine these similarities or differences. If symbolic number processing is rooted in non-symbolic numerosity processing, then the processing of the non-symbolic stimulus should not be affected by the format of the stimulus to which it needs to be compared.

Whereas similar behavioral distance effects were found for all tasks including non-symbolic stimuli, our ERP results show differences in the processing of the primes between the purely non-symbolic task in which two dot patterns were presented and the mapping task with non-symbolic primes (dot patterns) and symbolic targets (digits). Moreover, differences between the targets in the purely non-symbolic task and the mapping task with symbolic primes and non-symbolic targets were found. Processing of non-symbolic numerosity is thus affected by task, which is highly unlikely in the light of the ANS mapping account. However, the results could possibly still support the account, if the ERPs in the mapping tasks would show highly similar, but slightly delayed waveforms in the mapping task as compared to the non-symbolic task. Visual inspection of the waveforms does not support this. Instead, differences seem to occur mainly in amplitude instead of latency. For the non-symbolic primes, the amplitude in the mapping task was more positive between 125 and 400 ms than in the purely non-symbolic task on the anterior electrodes. For the targets, the amplitude was more positive for the mapping task as compared to the purely non-symbolic task between 115 and 275 ms and more positive for the purely non-symbolic task than the mapping task between 578 and 750 ms. These differences both early and late in the



processing stream suggest that different cognitive processes take place in the different tasks. As such, the data do not support the ANS mapping account. It suggests that symbolic stimuli are not intuitively mapped onto their non-symbolic counterparts, even when the task requires mapping. This is in line with previous research on mapping (Lyons et al., 2012) and studies showing a lack of correlation between distance effects in non-symbolic and symbolic tasks (Holloway and Ansari, 2009; Sasanguie et al., 2017).

A recent alternative to the ANS mapping account is symbolic processing based on symbol–symbol associations (Reynvoet and Sasanguie, 2016). This account suggests that whereas small symbolic numbers initially acquire meaning through mapping, larger symbolic numbers are learned through associations between symbolic numbers, such as “order” and “the successor function” (Carey, 2001, 2004, 2009). In adulthood, symbolic and non-symbolic numerosities would be processed independent from each other if tasks do not require relating them to each other (Lyons et al., 2012; Sasanguie et al., 2017). Both our behavioral and ERP data support this idea, as shown by the differences in ERPs between the tasks. However, the account on symbol–symbol associations does not directly lead to any predictions for mapping tasks.

In mapping tasks, contrary to what was proposed in the ANS mapping account, it may be the case that non-symbolic numerosities are first estimated and then compared to the symbolic number based on the symbol–symbol account. This may also explain the differences between the processing of the non-symbolic stimuli in the non-symbolic task vs. the mapping tasks. If a non-symbolic numerosity needs to be compared to a symbolic number, then it may first need to be estimated. However, if a non-symbolic numerosity needs to be compared to another non-symbolic numerosity, this is not necessary, which

is in line with the differences we found in the ERPs. This is also supported by research showing longer reaction times in mapping tasks involving symbolic and non-symbolic stimuli (Lyons et al., 2012). Additional support for this claim would come from similar processing of symbolic stimuli in the symbolic and mapping tasks. However, our paradigm does not allow to test this hypothesis, since the symbolic task did not require participants to process quantity at all. Since the same format (digits) was used for the primes and the target, the task could be performed by visual matching instead of matching based on quantity. Therefore, neither the ERPs nor the behavioral data give insight into the processing of symbolic number. Future research should include a different symbolic task, for example with number words and digits, to make sure participants process the numerical magnitude of the stimulus.

Taken together, our results support the converging evidence against the ANS theory and the ANS mapping account (Gebuis et al., 2016; Reynvoet and Sasanguie, 2016; Leibovich et al., 2017; Núñez, 2017). However, our lack of distance effects was based on null results. Whereas the analyses on the visual features with the same power did produce statistically significant results, the conclusions need to be interpreted with some caution. Research with a different paradigm showing similar results would strengthen our conclusions. For now, the results in the non-symbolic task do support the sensory-integration theory for processing non-symbolic numerosity (Gebuis et al., 2016) or sense for magnitude theory (Leibovich et al., 2017) instead. We suggest that mapping may be a two-step process, consisting of dot enumeration followed by comparison based on symbol–symbol associations (Reynvoet and Sasanguie, 2016). Future research including mapping tasks with purely symbolic stimuli, such as number words and Arabic numbers may shed further light on this issue.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the ethics committee of the Faculty of Social and Behavioral Sciences of the University of Utrecht. The protocol was approved by the ethics committee of the Faculty of Social and Behavioral Sciences of the University of Utrecht. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

All authors listed have made substantial, direct, and intellectual contributions to the work, and approved for publication. AvH and EK construed the study together. AvH gathered the data together with research assistants and master students. Analyses

were carried out by AvH. AvH and EK wrote the paper together.

FUNDING

This research was funded by the Dutch Scientific Organization (NWO), Aspasia grant number 015.008.028, awarded to EK.

ACKNOWLEDGMENTS

We would like to thank Dennis Hofman and Jos Jaspers for the technical support. We would also like to thank the Center for Information Technology of the University of Groningen for providing access to the Peregrine high performance computing cluster.

REFERENCES

- Carey, S. (2001). Cognitive foundations of arithmetic: evolution and ontogenesis. *Mind Lang.* 16, 37–55. doi: 10.1111/1468-0017.00155
- Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus* 133, 59–68. doi: 10.1162/001152604772746701
- Carey, S. (2009). Where our number concepts come from. *J. Philos.* 106, 220–254. doi: 10.5840/jphil2009106418
- Cohen Kadosh, R., and Walsh, V. (2009). Numerical representation in the parietal lobes: abstract or not abstract? *Behav. Brain Sci.* 32, 313–328. doi: 10.1017/S0140525X09990938
- Content, A., Vande Velde, M., and Adriano, A. (2017). Approximate number sense theory or approximate theory of magnitude? *Behav. Brain Sci.* 40:e168. doi: 10.1017/S0140525X16002089
- de Hevia, M. D., Castaldi, E., Streri, A., Eger, E., and Izard, V. (2017). Perceiving numerosity from birth. *Behav. Brain Sci.* 40:e169. doi: 10.1017/S0140525X16002090
- Dehaene, S. (1996). The organization of brain activations in number comparison: event-related potentials and the additive-factors method. *J. Cogn. Neurosci.* 8, 47–68. doi: 10.1162/jocn.1996.8.1.47
- Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. New York, NY: Oxford University Press.
- Dehaene, S., and Changeux, J. (1993). Development of elementary numerical abilities: a neuronal model. *J. Cogn. Neurosci.* 5, 390–407. doi: 10.1162/jocn.1993.5.4.390
- Dehaene, S., Dupoux, E., and Mehler, J. (1990). Is numerical comparison digital? analogical and symbolic effects in two-digit number comparison. *J. Exp. Psychol. Hum. Percept. Perform.* 16, 626–641. doi: 10.1037/0096-1523.16.3.626
- Dehaene, S., Izard, V., and Piazza, M. (2005). *Control Over Non-Numerical Parameters in Numerosity Experiments*. Available at: <http://www.unicog.org/docs/DocumentationDotsGeneration.doc>
- Gebuis, T., Cohen Kadosh, R., and Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: a critical review. *Acta Psychol.* 171, 17–35. doi: 10.1016/j.actpsy.2016.09.003
- Gebuis, T., and Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behav. Res. Methods* 43, 981–986. doi: 10.3758/s13428-011-0097-5
- Gebuis, T., and Reynvoet, B. (2012). Continuous visual properties explain neural responses to nonsymbolic number. *Psychophysiology* 49, 1481–1491. doi: 10.1111/j.1469-8986.2012.01461.x
- Gebuis, T., and Reynvoet, B. (2013). The neural mechanisms underlying passive and active processing of numerosity. *Neuroimage* 70, 301–307. doi: 10.1016/j.neuroimage.2012.12.048
- Gevers, W., Cohen Kadosh, R., and Gebuis, T. (2016). “Sensory integration theory: an alternative to the approximate number system,” in *Continuous Issues in Numerical Cognition*, ed. A. Henik (New York, NY: Elsevier), 405–418. doi: 10.1016/B978-0-12-801637-4.00018-4
- Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11116–11120. doi: 10.1073/pnas.1200196109
- Holloway, I. D., and Ansari, D. (2008). Domain-specific and domain-general changes in children’s development of number comparison. *Dev. Sci.* 11, 644–649. doi: 10.1111/j.1467-7687.2008.00712.x
- Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children’s mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001
- Hyde, D. C., and Spelke, E. S. (2009). All numbers are not equal: an electrophysiological investigation of small and large number representations. *J. Cogn. Neurosci.* 21, 1039–1053. doi: 10.1162/jocn.2009.21090
- Hyde, D. C., and Spelke, E. S. (2012). Spatiotemporal dynamics of processing nonsymbolic number: an event-related potential source localization study. *Hum. Brain Mapp.* 33, 2189–2203. doi: 10.1002/hbm.21352
- Izard, V., Sann, C., Spelke, E. S., and Streri, A. (2009). Newborn infants perceive abstract numbers. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10382–10385. doi: 10.1073/pnas.0812142106
- Krajcsi, A., Lengyel, G., and Kojouharova, P. (2018). Symbolic number comparison is not processed by the analogue number system: different symbolic and nonsymbolic numerical distance and size effects. *Front. Psychol.* 9:124. doi: 10.3389/fpsyg.2018.00124
- Leibovich, T., Katzin, N., Harel, M., and Henik, A. (2017). From “sense of number” to “sense of magnitude”: the role of continuous magnitudes in numerical cognition. *Behav. Brain Sci.* 40:e164. doi: 10.1017/S0140525X16000960
- Libertus, M. E., Braham, E. J., and Liu, R. (2017). Infants discriminate number: evidence against the prerequisite of visual object individuation and the primacy of continuous magnitude. *Behav. Brain Sci.* 40:e176. doi: 10.1017/S0140525X16002144
- Libertus, M. E., Woldorff, M. G., and Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. *Behav. Brain Funct.* 3:1.
- Loourenco, S. F., Ayzenberg, V., and Lyu, J. (2016). A general magnitude system in human adults: evidence from a subliminal priming paradigm. *Cortex* 81, 93–103. doi: 10.1016/j.cortex.2016.04.013
- Lyons, I. M., Ansari, D., and Beilock, S. L. (2012). Symbolic estrangement: evidence against a strong association between numerical symbols and the quantities they represent. *J. Exp. Psychol. Gen.* 141, 635–641. doi: 10.1037/a0027248
- Mix, K. S., Huttenlocher, J., and Levine, S. C. (2002). Multiple cues for quantification in infancy: is number one of them? *Psychol. Bull.* 128, 278–294. doi: 10.1037/0033-2909.128.2.278
- Moyer, R. S., and Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature* 215, 1519–1520. doi: 10.1038/2151519a0

- Nieder, A. (2017). Number faculty is alive and kicking: on number discriminations and number neurons. *Behav. Brain Sci.* 40:e181. doi: 10.1017/S0140525X16002193
- Núñez, R. E. (2017). Is there really an evolved capacity for number? *Trends Cogn. Sci.* 21, 409–424. doi: 10.1016/j.tics.2017.03.005
- Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011:156869. doi: 10.1155/2011/156869
- Park, J., DeWind, N. K., and Brannon, E. M. (2017). Direct and rapid encoding of numerosity in the visual stream. *Behav. Brain Sci.* 40:e185. doi: 10.1017/S0140525X16002235
- Paulsen, D. J., and Neville, H. J. (2008). The processing of non-symbolic numerical magnitudes as indexed by ERPs. *Neuropsychologia* 46, 2532–2544. doi: 10.1016/j.neuropsychologia.2008.04.003
- Reynvoet, B., and Sasanguie, D. (2016). The symbol grounding problem revisited: a thorough evaluation of the ANS mapping account and the proposal of an alternative account based on Symbol–Symbol associations. *Front. Psychol.* 7:1581. doi: 10.3389/fpsyg.2016.01581
- Sasanguie, D., De Smedt, B., Defever, E., and Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *Br. J. Dev. Psychol.* 30, 344–357. doi: 10.1111/j.2044-835X.2011.02048.x
- Sasanguie, D., De Smedt, B., and Reynvoet, B. (2017). Evidence for distinct magnitude systems for symbolic and non-symbolic number. *Psychol. Res.* 81, 231–242. doi: 10.1007/s00426-015-0734-1
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: what underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418–431. doi: 10.1016/j.jecp.2012.10.012
- Savelkoul, S., and Cordes, S. (2017). Numerical intuitions in infancy: give credit where credit is due. *Behav. Brain Sci.* 40:e189. doi: 10.1017/S0140525X16002272
- Smets, K., Moors, P., and Reynvoet, B. (2016). Effects of presentation type and visual control in numerosity discrimination: implications for number processing? *Front. Psychol.* 7:66. doi: 10.3389/fpsyg.2016.00066
- Smets, K., Sasanguie, D., Szűcs, D., and Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *J. Cogn. Psychol.* 27, 310–325. doi: 10.1080/20445911.2014.996568
- Soltész, F., and Szűcs, D. (2014). Neural adaptation to non-symbolic number and visual shape: an electrophysiological study. *Biol. Psychol.* 103, 203–211. doi: 10.1016/j.biopsycho.2014.09.006
- Stoianov, I. P., and Zorzi, M. (2017). Computational foundations of the visual number sense. *Behav. Brain Sci.* 2017, 1–62. doi: 10.1017/S0140525X16002326
- Temple, E., and Posner, M. I. (1998). Brain mechanisms of quantity are similar in 5-year-old children and adults. *Proc. Natl. Acad. Sci. U.S.A.* 95, 7836–7841. doi: 10.1073/pnas.95.13.7836
- Van Opstal, F., Gevers, W., De Moor, W., and Verguts, T. (2008). Dissecting the symbolic distance effect: comparison and priming effects in numerical and nonnumerical orders. *Psychon. Bull. Rev.* 15, 419–425. doi: 10.3758/PBR.15.2.419
- van Opstal, F., and Verguts, T. (2011). The origins of the numerical distance effect: the same–different task. *J. Cogn. Psychol.* 23, 112–120. doi: 10.1080/20445911.2011.466796
- Verguts, T., and Van Opstal, F. (2005). Dissociation of the distance effect and size effect in one-digit numbers. *Psychon. Bull. Rev.* 12, 925–930. doi: 10.3758/BF03196787
- Xu, F., and Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition* 74, B1–B11. doi: 10.1016/S0010-0277(99)00066-9
- Xu, F., Spelke, E. S., and Goddard, S. (2005). Number sense in human infants. *Dev. Sci.* 8, 88–101. doi: 10.1111/j.1467-7687.2005.00395.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 van Hoogmoed and Kroesbergen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Using Hierarchical Linear Models to Examine Approximate Number System Acuity: The Role of Trial-Level and Participant-Level Characteristics

Emily J. Braham^{1,2*}, Leanne Elliott^{1†} and Melissa E. Libertus^{1,2}

¹ Department of Psychology, University of Pittsburgh, Pittsburgh, PA, United States, ² Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA, United States

OPEN ACCESS

Edited by:

Xinlin Zhou,
Beijing Normal University, China

Reviewed by:

Nicholas Kurshan DeWind,
University of Pennsylvania,
United States
Veronica Mazza,
University of Trento, Italy

*Correspondence:

Emily J. Braham
ejb67@pitt.edu

[†] These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 15 May 2018

Accepted: 09 October 2018

Published: 12 November 2018

Citation:

Braham EJ, Elliott L and
Libertus ME (2018) Using Hierarchical
Linear Models to Examine
Approximate Number System Acuity:
The Role of Trial-Level
and Participant-Level Characteristics.
Front. Psychol. 9:2081.
doi: 10.3389/fpsyg.2018.02081

The ability to intuitively and quickly compare the number of items in collections without counting is thought to rely on the Approximate Number System (ANS). To assess individual differences in the precision of peoples' ANS representations, researchers often use non-symbolic number comparison tasks in which participants quickly choose the numerically larger of two arrays of dots. However, some researchers debate whether this task actually measures the ability to discriminate approximate numbers or instead measures the ability to discriminate other continuous magnitude dimensions that are often confounded with number (e.g., the total surface area of the dots or the convex hull of the dot arrays). In this study, we used hierarchical linear models (HLMs) to predict 132 adults' accuracy on each trial of a non-symbolic number comparison task from a comprehensive set of trial-level characteristics (including numerosity ratio, surface area, convex hull, and temporal and spatial variations in presentation format) and participant-level controls (including cognitive abilities such as visual-short term memory, working memory, and math ability) in order to gain a more nuanced understanding of how individuals complete this task. Our results indicate that certain trial-level characteristics of the dot arrays contribute to our ability to compare numerosities, yet numerosity ratio, the critical marker of the ANS, remains a highly significant predictor of accuracy above and beyond trial-level characteristics and across individuals with varying levels of math ability and domain-general cognitive abilities.

Keywords: approximate number system, numerosity, math ability, surface area, convex hull, hierarchical linear model

INTRODUCTION

Without the use of symbols, counting, or formal mathematics, adults are able to rapidly estimate and compare the number of items in collections; we choose the bag of apples at the grocery store that contains the most apples, choose the parking lot that has the fewest cars, and stand in the check-out line that appears to have the fewest people. According to some researchers, the ability to intuitively compare approximate quantities taps into the Approximate Number System (ANS),

a system in which we process numbers as noisy or imprecise magnitudes with overlap between neighboring representations of number (Dehaene, 1992; Barth et al., 2006). In the ANS, the degree of overlap between neighboring quantity representations increases for larger quantities and the discriminability between two numbers is determined by the numerical ratio between them. For example, quickly approximating if a bag with 11 apples has more than a bag with 10 apples is more difficult than quickly approximating if a bag with 11 apples has more than a bag with 7 apples. In addition, determining that 11 apples are more than 7 apples is as easy as determining that 22 apples are more than 14 apples. Thus, the critical marking of ANS processing is ratio-dependent performance (Dehaene, 1992).

To assess the acuity of children's and adults' ANS representations, researchers most frequently use non-symbolic number comparison tasks in which participants quickly choose the numerically larger of two arrays of dots over a series of trials that vary in the difficulty of the ratio between the two arrays. Across variations in temporal and spatial characteristics of the stimulus presentation, participants are generally faster and more accurate with relatively more disparate numerosities compared to less disparate ones (Dehaene, 1992; Cantlon and Brannon, 2006; Libertus et al., 2007; Halberda and Feigenson, 2008; Halberda et al., 2008; Soltész et al., 2010; Inglis et al., 2011; Dewind and Brannon, 2012; Price et al., 2012; Agrillo et al., 2013).

However, some researchers debate whether tasks designed to measure approximate number discrimination instead measure the ability to discriminate other perceptual variables that are confounded with number (Gebuis and Reynvoet, 2012; Leibovich et al., 2016; Henik et al., 2017). Here, we apply a novel analysis method, namely hierarchical linear modeling (HLM), to predict individual participants' accuracy on each trial of a non-symbolic number comparison task from multiple trial-level characteristics (perceptual variables, presentation format) and participant-level controls (i.e., cognitive abilities such as visual-short term memory, working memory, and math ability) that are likely linked to performance on non-symbolic number comparison tasks. These analyses allow for greater specificity in unpacking the influence of several confounds simultaneously to account for differences in performance on the task both within and between individuals.

The Role of Perceptual Variables for Non-symbolic Number Comparisons

In everyday life, number is frequently correlated with other visual characteristics (e.g., more apples take up more space). In non-symbolic number comparison tasks, non-numeric continuous dimensions of the dot arrays, such as cumulative area, cumulative perimeter, dot size, and/or visual density can influence judgments about numerosity (e.g., Allik and Tuulmets, 1991; Durgin, 1995; Tokita and Ishiguchi, 2010; Dewind and Brannon, 2012). Researchers often attempt to rule out the use of these non-numeric continuous dimensions such that they are not consistently confounded with number throughout the entire experiment. However, these methods have been

criticized for only manipulating a small subset of continuous magnitudes in any given trial, and thus allowing participants to use the other non-manipulated continuous magnitudes to predict numerosity (Gebuis and Reynvoet, 2012). For example, participants may use non-numerical visual cues such as convex hull or density to make numerosity judgments even when other visual features such as cumulative surface are not confounded with numerosity. Others have criticized this approach for not carefully accounting for all continuous dimensions (Clayton et al., 2015; Gilmore et al., 2016). For example, images from the freely available Panamath software¹ are frequently used in the literature (Halberda and Feigenson, 2008; Halberda et al., 2008; Libertus et al., 2011, 2013a,b; Mazzocco et al., 2011; Libertus et al., 2012; Fazio et al., 2014; Hyde et al., 2014; van Marle et al., 2014; Haist et al., 2015; Norris et al., 2015; Patalano et al., 2015; Purpura and Logan, 2015; Bugden and Ansari, 2016; Norris and Castronovo, 2016; Braham and Libertus, 2017, 2018; Dillon et al., 2017; Lukowski et al., 2017; Geary et al., 2018), yet the software does not allow researchers to manipulate convex hull (i.e., the area of the smallest polygon that encompasses all of the dots in the set). Studies have demonstrated that convex hull is confounded with number in Panamath images, such that the more numerous set in each image typically also has a larger convex hull (Clayton et al., 2015; DeWind and Brannon, 2016). In a recent study, Gilmore et al. (2016) compared the influence of convex hull and cumulative surface area (which was highly correlated with dot diameter and density of the array) on both children's and adults' numerosity judgments on a non-symbolic comparison task. Convex hull information influenced accuracy across all age groups such that children and adults were more accurate on number comparisons when the convex hull ratio was large, but cumulative surface area information only influenced children's, and not adults', accuracy on number comparisons. These findings suggest that it is more difficult for adults to ignore convex hull information compared to cumulative surface area information.

Recent studies have used a new approach to constructing dot arrays that involves intentionally and systematically varying numerosity and non-numerical continuous dimensions in relation to one another in order to disentangle their influence on numerosity judgments (DeWind et al., 2015; DeWind and Brannon, 2016; Park et al., 2016; Starr et al., 2017). In these stimuli, features of the dot arrays are reduced to three parameters: number, size (i.e., the features related to individual element size, total surface area, and total perimeter), and spacing [i.e., the features related to convex hull and sparsity (convex hull/number of items)]. Using a modeling approach, DeWind et al. (2015) were able to dissociate the influence of the size and spacing features and show that while size and spacing bias adults' numerosity judgments, the effect of these features was relatively small. Both children and adults primarily use number in numerical discrimination tasks, rather than size or spacing (DeWind et al., 2015; Starr et al., 2017). Further, there is evidence for earlier neural sensitivity to numerosity

¹www.panamath.org

compared to these other continuous dimension features (Park et al., 2016).

The Role of Spatial and Temporal Presentation Format for Non-symbolic Number Comparisons

Across studies that use non-symbolic number comparison tasks, there is also wide variation in the presentation format of the dot displays; some studies present the two arrays of dots simultaneously side-by-side, with spatial separation (i.e., one on either side of the screen or paired presentation), while other studies simultaneously present two arrays of different colors with spatial overlap (i.e., intermixed presentation). Most studies in the literature exclusively use either separated displays (e.g., Halberda and Feigenson, 2008; Piazza et al., 2010; Inglis et al., 2011; Lyons and Beilock, 2011; Libertus et al., 2012; Gilmore et al., 2013) or exclusively use overlapping displays (e.g., Dewind and Brannon, 2012; Halberda et al., 2012; Lourenco et al., 2012; Lindskog et al., 2013), with only a few studies using both presentation formats (Price et al., 2012; Norris and Castronovo, 2016). In a recent study using Panamath images, Norris and Castronovo (2016) directly compared different groups of participants' accuracy on non-symbolic number comparison tasks using either spatially separated or spatially overlapping displays. Accuracy was higher and more reliable for participants who viewed the spatially separated displays compared to the overlapping displays. Lower performance on spatially overlapping displays may reflect the additional cognitive processing required to visually segment the arrays (Price et al., 2012).

A second major distinction in format across studies lies in the temporal aspects of the presentation. In the studies described above researchers presented the two spatially separated or spatially overlapping dot arrays simultaneously; however, a number of studies instead display the dot arrays sequentially, with one array followed by the other (Ansari et al., 2007; Hayashi et al., 2013). Smets et al. (2016) used a within-subjects design to directly compare participants' performance on simultaneous trials, presented for 1500 ms, and sequential trials, in which each array was presented for 750 ms with a 500-ms pause between arrays. Participants had overall higher accuracy when arrays were presented simultaneously than when they were presented sequentially. There are a few potential explanations for these results. First, it has been suggested that additional working memory resources are required when the arrays of dots are presented successively (Price et al., 2012). Second, simultaneously presented side-by-side arrays may allow for more fine-grained, explicit comparisons of the two arrays than is possible on sequential trials in which only the second array can be kept in visual-spatial short-term memory (Brown and Rebbin, 1970; Smets et al., 2014, 2016). Thus, when images are presented sequentially, participants may use an alternative strategy in which they extract the numerosity of the first array to compare it to the numerosity of the second array (Frick, 1985; Smets et al., 2016).

These methodological differences in the spatial and temporal aspects of the dot displays are clearly present across studies

yet infrequently accounted for in the literature. To our knowledge, only one study to date included all three presentation formats described above (simultaneously presented with spatial separation, simultaneously presented with spatial overlap, and sequentially presented) within a single study (Price et al., 2012). In a within-subjects design, Price et al. (2012) found significant positive correlations between participants' performance in all formats of the task. In line with the findings of Norris and Castronovo (2016), participants' performance was significantly worse on the simultaneously presented, spatially overlapping trials compared to the other two types of trials. However, unlike the results of Smets et al. (2016), there was no difference in participants' performance on the simultaneously presented, spatially separated trials compared to the sequential trials. It is important to note that performance was measured using Weber fractions—an index of the imprecision of participants' ANS representations—which has been shown to be a less reliable measure of ANS acuity compared to accuracy (Inglis and Gilmore, 2014). Nevertheless, together these findings suggest that performance on non-symbolic number comparison tasks is not independent of the spatial and temporal aspects of the presentation and that differences in accuracy across formats may be due to extraneous domain-general cognitive demands.

The Link Between Non-symbolic Number Comparison Performance and Math Ability

Many studies propose a link between performance on non-symbolic number comparison tasks and measures of math ability, which involve using exact or symbolic representations of numbers to count and perform exact calculations (Halberda and Feigenson, 2008; Halberda et al., 2008, 2012; Gilmore et al., 2010; Inglis et al., 2011; Mazzocco et al., 2011; Libertus et al., 2011, 2012, 2013b; Dewind and Brannon, 2012; Lourenco et al., 2012; Bonny and Lourenco, 2013; Guillaume et al., 2013; Keller and Libertus, 2015; Braham and Libertus, 2017, 2018). These studies offer several potential explanations for the relation between the ANS and math. First, when children acquire knowledge of new symbolic numbers, they may map their new symbolic representations to their existing underlying ANS representations (Brankaer et al., 2014; Pinheiro-Chagas et al., 2014). Second, an intuitive understanding of approximate arithmetic with non-symbolic quantities may serve as a foundation for understanding symbolic arithmetic (Park and Brannon, 2014; Pinheiro-Chagas et al., 2014). And third, ANS representations may help facilitate error detection, as people with more precise ANS representations may more easily notice magnitude errors when performing symbolic calculations on a math assessment (Lourenco et al., 2012; Feigenson et al., 2013).

Although a number of meta-analyses provide support for the correlation between ANS acuity and math ability (Chen and Li, 2014; Fazio et al., 2014; Schneider et al., 2016), the correlations are overall low or moderate and there are many studies that report null or mixed results (Holloway and Ansari, 2009; Soltész et al., 2010; Castronovo and Göbel, 2012;

Price et al., 2012; Fuhs and McNeil, 2013; Kolkman et al., 2013; Sasanguie et al., 2013). The discrepancy in findings across studies may be partly due to methodological differences in the way that math skills are assessed (Schneider et al., 2016; Braham and Libertus, 2018) or the way the non-symbolic number comparison task is constructed, including the spatial and temporal aspects of the presentation format and the controls for non-numerical continuous dimensions of the dot arrays (Norris and Castronovo, 2016). The inconsistent relation between ANS acuity and math ability across studies may also relate to participant-level characteristics of the sample, such as age (Inglis et al., 2011), individual differences in domain general cognitive skills that are needed across both tasks (e.g., working memory or inhibitory control; Fuhs and McNeil, 2013; Gilmore et al., 2013; Keller and Libertus, 2015), or other characteristics of the participants that often go unmeasured in these studies (e.g., math anxiety; Lindskog et al., 2017; Braham and Libertus, 2018).

The Current Study

Although several studies have explored how specific trial-level characteristics, such as continuous magnitude dimensions or spatial and temporal presentation format, influence participants' accuracy on non-symbolic number comparison tasks, less is known about how these variables operate uniquely from one another and potentially modulate numerosity ratio effects, a hallmark of non-symbolic numerical processing. In the present study, we used HLMs to predict people's accuracy on the non-symbolic number comparison task from a comprehensive set of trial-level characteristics and participant-level controls. An advantage of this modeling approach is that it allows for the simultaneous estimation of the variation from person to person as well as from trial to trial. Here, we use a single model to simultaneously examine which features of the dot stimuli and which aspects of domain-general cognition relate to non-symbolic number comparison performance. We specifically address the following three research questions. First, how do trial-level characteristics, including numerosity ratio, spatial and temporal aspects of the presentation format, and continuous magnitude dimensions, and participant-level characteristics, including age, gender, math ability, phonological working memory, and visuospatial short-term memory, uniquely and independently relate to performance on individual trials of a non-symbolic number comparison task? Here we specifically focus on two continuous magnitude dimensions, cumulative surface area and convex hull, which are independent of each other and have been identified in the literature as potentially confounding variables (Gebuis and Reynvoet, 2012; DeWind and Brannon, 2016). As a robustness check, we also estimate these models with measures of average dot area and density included in the place of cumulative surface area and convex hull. Second, to what extent do these trial-level characteristics moderate the association between numerosity ratio and accuracy? Finally, to what extent does math ability moderate associations between these trial-level characteristics and individual's accuracy on the non-symbolic number comparison task?

MATERIALS AND METHODS

Participants

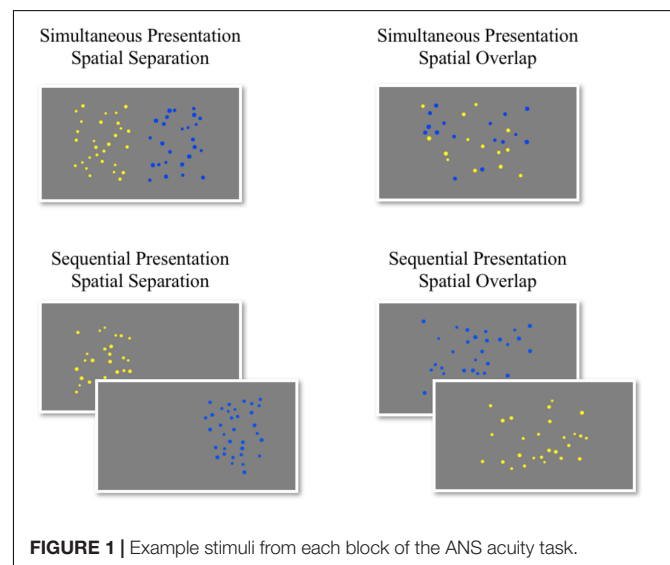
One-hundred thirty-five undergraduate students participated in a laboratory study in exchange for course credit. Three participants were excluded from all analyses due to incomplete data: two participants did not complete all measures of working memory and one participant did not report their gender. The final sample consisted of 132 participants (69 males) who ranged in age from 18 to 52 years of age ($M = 19.71$; $SD = 4.23$). The majority of our participants were in their first year of university ($n = 83$) and identified their race as White ($n = 103$). A subset of this sample completed a more extensive battery of tasks and those data have been previously reported elsewhere (Braham and Libertus, 2018).

Measures

ANS Acuity

To measure ANS acuity, participants completed a total of 360 trials of a non-symbolic number comparison task in which they were presented with arrays of blue and yellow dots on a computer screen and instructed to select the color with more dots as quickly and accurately as possible. On all trials, participants indicated their response by pressing one of two keys on the keyboard, marked with either a yellow or a blue sticker. The correct response (i.e., the color with more dots) was counterbalanced across trials and participants received trial-level feedback—they heard a beep if they responded incorrectly.

The 360 trials were divided into four blocks (90 trials per block) that varied in the spatial (spatial separation vs. spatial overlap) and temporal aspects (simultaneous vs. sequential presentation) of the stimulus presentation in an orthogonal design: (1) simultaneous presentation with spatial separation, (2) simultaneous presentation with spatial overlap, (3) sequential presentation with spatial separation, and (4) sequential presentation with spatial overlap (Figure 1). Participants completed the blocks in a counterbalanced order.



All trials started with a fixation cross for 500 ms. On blocks with simultaneous presentation of the arrays, the blue and yellow dots appeared for 1500 ms; on blocks with sequential presentation of the arrays, one array appeared for 750 ms followed by the other for 750 ms. Participants could select their response on the keyboard either during the display of the dot arrays or during the blank screen that followed. Three participants were missing one ($n = 2$) or two ($n = 1$) blocks of this task but were retained in the analyses.

The images were presented using a custom-made Matlab script. All stimuli were extracted from the Psychological Assessment of Numerical Ability (Panamath)². Each dot array contained between 12 and 36 dots and appeared on a gray background. Dot size varied within single arrays (average dot diameter = 36 pixels; allowed variation = 20%). The ratio of the larger quantity of dots to the smaller quantity of dots was evenly split across trials in one of five numerosity ratio categories (72 trials per ratio): 1.11, 1.14, 1.2, 1.25, 1.33. Surface area and convex hull ratios were calculated by dividing the value from the more numerous array by the value from the less numerous array. Surface area ratios ranged from 0.72 to 1.35. Convex hull ratios ranged from 0.72 to 1.71.

Math Ability

Participants' math abilities were assessed using the Math Fluency subtest of the Woodcock Johnson III Tests of Achievement (Woodcock et al., 2001). Participants were presented with 160 simple addition, subtraction, and multiplication problems containing numbers in the 1–10 range (e.g., $8 - 0 = _$; $3 \times 6 = _$). They were told to begin with the first problem, to work quickly and accurately, and to solve as many problems as they could within the 3-min time limit. The raw score (number of problems solved correctly) was converted into an age-normed standardized score with an expected mean of 100 and standard deviation of 15.

Visuospatial Short-Term Memory

We used a computerized flicker change detection task to assess participants' visuospatial short-term memory capacity (Pailian and Halberda, 2015). On each trial, participants were presented with two arrays of yellow and blue dots on a gray background in continuous alteration. Each array flashed on the computer screen for 700 ms with a 900-ms pause between arrays. The two arrays were identical except for the color of one dot. Participants were told to search for the "target" dot (i.e., the dot that changed in color between the two images) as quickly and accurately as possible. They were instructed to press the space bar on the keyboard as soon as they detected the target to record their response time and to freeze the display, and then to use the computer mouse to click on the target dot to record their response. There were a total of 90 trials and the set size of the displays was manipulated across trials: 1/3 of the trials contained arrays with 6 dots, 1/3 of the trials contained arrays with 8 dots, and 1/3 of the trials contained arrays with 10 dots. Average response time on the correct trials, excluding trials in which

participants' response times were over two standard deviations from their average trial response time, was used as the measure of participants' visual short-term memory with longer response times indicating smaller visual short-term memory capacity.

Phonological Working Memory

To assess phonological working memory, participants completed a backward digit span task, in which they listened to series of digit sequences presented at a rate of one item per second (e.g., "5, 9, 1, 3, 7") and were instructed to recall the sequence in reverse order (e.g., "7, 3, 1, 9, 5"). The length of the sequences increased in difficulty throughout the task from three digits to 12 digits and participants were presented with two trials for each sequence length. Participant responses were marked as either correct or incorrect. Administration continued until the participant gave incorrect responses to both trials of the same sequence length. The length of the longest sequence in which the participant recalled at least one of the trials correctly was used as the participants' phonological working memory span score.

Procedure

All participants provided written, informed consent prior to participation. The study took place in a quiet laboratory room during a single 1-h session. Participants completed the tasks in the following order: ANS acuity, visuospatial short-term memory, phonological working memory, math fluency.

Analysis Plan

A series of 2-level logistic hierarchical linear models (HLMs) were estimated to predict individual participants' accuracy on each trial of the non-symbolic number comparison task (47,160 observations). These models predict accuracy on each trial of the task (1 = correct, 0 = incorrect). Trial-level characteristics, including numerosity ratio, surface area ratio, convex hull ratio, spatial presentation format (i.e., spatially separated vs. overlapping), and temporal presentation format (i.e., simultaneous vs. sequential) were included as level-1 predictors. Participant-level characteristics, including math fluency, age, gender, phonological working memory, and visuospatial short-term memory were entered at level-2 as predictors of level-1 intercept (i.e., individual's average accuracy). Random intercepts by participant were included to account for individual differences in participants' average accuracy across all trials. Descriptive statistics for all study variables, including trial-level characteristics as well as participant-level characteristics, are shown in **Table 1**.

First, main effects of trial-level characteristics on accuracy were estimated, controlling for participant-level characteristics. Surface area and convex hull ratios were natural log transformed, such that a surface area or convex hull ratio of 0 indicates that the continuous magnitude is equated across sets (as the untransformed ratio would be equal to 1), negative values indicate that the less numerous array had a larger value of this continuous magnitude, and positive values indicate that the more numerous array had a larger value of this continuous magnitude. Continuous indicators of surface area and convex hull were used in the analyses shown here as they offer more

²www.panamath.org

TABLE 1 | Descriptive statistics of level-1, trial-level characteristics ($N = 360$) and of level-2, participant-level characteristics ($N = 132$).

Level 1 Trial-Level Characteristics	<i>M</i> (<i>SD</i>)/%	Range
Numerosity Ratio (in raw values)	1.21 (0.08)	1.11, 1.33
Surface Area Ratio (in raw values)	1.01 (0.16)	0.73, 1.34
Convex Hull Ratio (in raw values)	1.12 (0.15)	0.83, 1.66
Dot Size Ratio (in raw values)	0.84 (0.14)	0.55, 1.02
Density Ratio (in raw values)	1.10 (0.15)	0.75, 1.61
Spatial Arrangement		
Overlapping	50%	
Separated	50%	
Presentation Format		
Sequential	50%	
Simultaneous	50%	
Level 2 Participant-Level Characteristics	<i>M</i> (<i>SD</i>)/%	
Average Trial Accuracy	0.76 (0.05)	0.63, 0.89
Math Fluency	105.45 (13.17)	65, 153
Age (in years)	19.71 (4.23)	18, 52
Female	48%	
Visuospatial Short-Term Memory	1.10 (0.67)	0.27, 3.27
Phonological Working Memory	4.94 (1.40)	2, 8

specificity regarding the degree to which continuous magnitudes are positively or negatively correlated with number³. Numerosity ratio was also centered at 1, such that a value of 0 indicates no difference in the two numbers (i.e., 1:1 ratio), and rescaled by a factor of 10, such that a one unit change in the rescaled variable represented a 0.1 unit change in ratio, for interpretability. Correlations among these transformed trial-level variables are shown in **Table 2**. All continuous level-2 variables were grand-mean centered.

To answer our second research question regarding whether trial-level characteristics moderate associations between numerical ratio and accuracy, a series of interactions were then tested between numerical ratio and each additional trial-level characteristic. Interactions were first entered individually,

³Trials could also be categorized as congruent (i.e., the array with the larger number had the larger cumulative area or convex hull), equated (i.e., the arrays had equal cumulative area or convex hull), and incongruent (i.e., the array with the smaller number had the larger cumulative area or convex hull). Models using these categorical indicators of congruency instead of the continuous ones yielded similar results to the ones described in the text.

and then all significant interactions were entered into a single model. Simple effects of numerical ratio predicting accuracy were then calculated at various levels of these moderating trial-level characteristics to probe significant interactions.

To answer our third research question regarding the role of math ability in these associations, we first included math ability as a level-2 predictor of level-1 intercept in order to address whether individuals with higher levels of math ability had higher overall accuracy on the non-symbolic number comparison task. Math ability was then included as a predictor of the ratio slope (i.e., as a cross-level interaction) to examine whether the magnitude of ratio effects differed across individuals with varying levels of math ability.

Finally, each of these models was estimated a second time with alternative measures of the perceptual variables described above. Specifically, raw cumulative surface area ratios were divided by the raw number ratio to represent the average dot size ratio of the larger set compared to the smaller set. Average dot size ratio ranged from 0.55 to 1.02. Additionally, raw number ratios were divided by the raw convex hull ratios to yield a ratio of the density of the larger set compared to the smaller set. Density ratio ranged from 0.75 to 1.61. Dot area ratio and density ratio were then natural log transformed and included as trial-level predictors in the place of surface area ratio and convex hull ratio respectively.

RESULTS

Main Effects of Trial-Level and Participant-Level Characteristics

Results of models estimating main effects of trial-level and participant-level characteristics on individuals' performance on the non-symbolic number comparison task are shown in the first column of **Table 3**. Numerosity ratio was a highly significant predictor of accuracy, as a 0.1 increase in numerical ratio (e.g., the difference between a 1.2 and 1.3 ratio) resulted in a 71% increase in the odds of correctly identifying the more numerous array. In other words, individuals were more accurate on trials in which the ratio of difference between the two arrays was larger, consistent with theoretical accounts of the ANS. Crucially, this association between numerosity ratio and accuracy was evident when controlling for continuous magnitude dimensions and

TABLE 2 | Correlations between trial-level characteristics for spatially separated and spatially overlapping trials.

	1	2	3	4	5
(1) Numerosity Ratio	—	−0.01	−0.37***	0.10	0.37***
(2) Surface Area Ratio	−0.001	—	0.93***	0.14	−0.13
(3) Dot Size Ratio	−0.37***	0.93***	—	0.09	−0.26*
(4) Convex Hull Ratio	−0.02	−0.07	0.07	—	−0.89***
(5) Density Ratio	0.53***	−0.06	−0.25*	−0.86***	—

Correlations above the diagonal describe trials in which dot arrays were spatially separated ($n = 90$), whereas correlations below the diagonal describe trial in which arrays were spatially overlapping ($n = 90$). Correlations are based on the log-transformed surface area, dot size, convex hull, and density values. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

TABLE 3 | Results of two-level logistic hierarchical linear models predicting trial-level accuracy on the non-symbolic number comparison task (1 = correct response) from trial-level and participant-level characteristics.

Fixed effects	OR (SE)	OR (SE)	OR (SE)
Trial-Level Characteristics			
Numerosity Ratio	1.71*** (0.03)	1.97*** (0.06)	1.98*** (0.06)
Surface Area Ratio	1.52*** (0.11)	3.30*** (0.80)	3.30*** (0.83)
Convex Hull Ratio	2.33*** (0.23)	2.98*** (0.87)	2.98*** (0.87)
Spatially Overlapping Presentation	0.66*** (0.02)	1.17* (0.08)	1.17** (0.08)
Sequential Presentation	1.12*** (0.02)	0.96 (0.06)	0.96 (0.06)
Participant-Level Characteristics			
Math Fluency	1.004* (0.002)	1.004* (0.002)	0.999 (0.003)
Age	0.99 (0.01)	0.99 (0.01)	0.99† (0.01)
Female	0.95 (0.05)	0.97 (0.05)	0.97 (0.05)
Visuospatial Short-Term Memory	1.003 (0.04)	1.003 (0.04)	1.003 (0.001)
Phonological Working Memory	1.02 (0.02)	1.02 (0.02)	1.02 (0.02)
Trial-Level Interactions			
Surface Area Ratio * Numerosity Ratio		0.73** (0.07)	0.73** (0.07)
Convex Hull Ratio * Numerosity Ratio		0.86 (0.12)	0.86 (0.12)
Spatially Overlapping Presentation * Numerosity Ratio		0.74*** (0.02)	0.74*** (0.02)
Sequential Presentation * Numerosity Ratio		1.09** (0.03)	1.09** (0.03)
Cross-Level Interactions			
Math Fluency * Numerosity Ratio			1.003* (0.001)
Intercept	1.17*** (0.06)	0.90 (0.06)	0.90 (0.06)
Random effect			
	Estimate (SE)	Estimate (SE)	Estimate (SE)
Intercept	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)

Values shown in the table are odds ratios and their standard errors. Numerosity ratio was centered at 1, and surface area ratio and convex hull ratio were natural log transformed. Math fluency scores were mean-centered prior to estimating models. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

variations in spatial and temporal presentation format of the task.

Furthermore, predicted accuracy significantly increased as surface area and convex hull ratios increased (i.e., as congruency between numerosity and surface area or convex hull increased). A one unit increase in convex hull congruency (i.e., the difference between trials in which convex hull was equal across sets, where this variable would be equal to 0, and trials in which the convex hull of the larger set was 2.72 times the size of the smaller set, where this variable would have a value of 1) resulted in a 133% increase in the odds of responding correctly, even when holding numerical ratio and other trial-level and participant-level characteristics constant. Similarly, a one unit increase in surface area ratio (i.e., the difference between trials in which cumulative surface was equal across sets, where this variable would be equal to 0, and trials in which the cumulative surface area of the larger set was 2.72 times the size of the smaller set, where this variable would have a value of 1) was associated with a 52% increase in the odds of responding correctly, controlling for numerical ratio and other trial-level and participant-level characteristics. Additionally, individuals tended to be more accurate on trials where arrays were presented with spatial separation (52% higher odds of responding correctly) and where arrays were presented sequentially (12% higher odds of correct response).

Few participant-level characteristics predicted level-1 intercepts at level-2. Math fluency scores were positively related

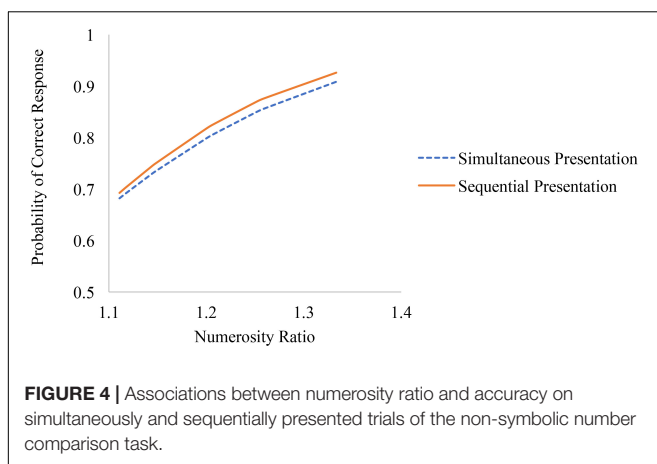
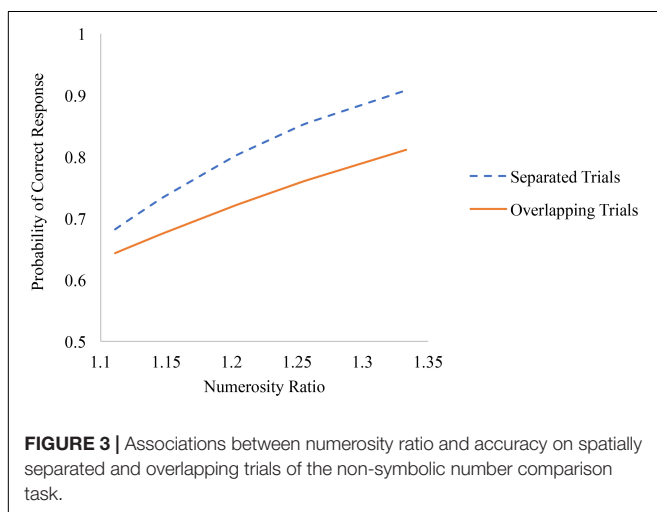
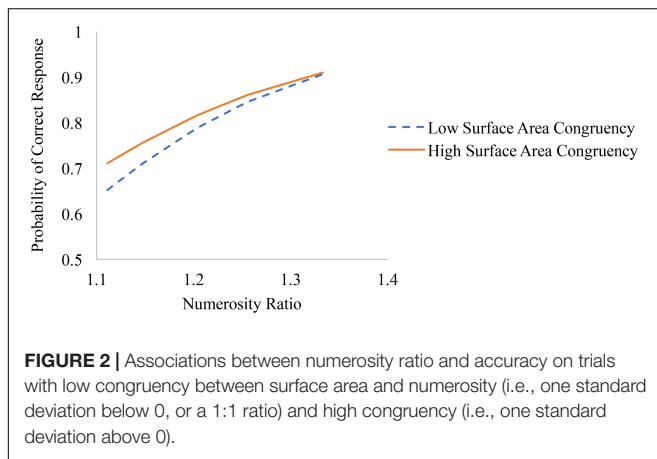
to overall accuracy, such that a standard deviation increase in math fluency predicted a 7% increase in odds ratio. However, participant age, gender, phonological working memory, and visuospatial short-term memory were unrelated to overall accuracy in these models.

Trial-Level Interactions With Numerosity Ratio

Interactions between trial-level characteristics and numerosity ratio were then entered into models individually. Surface area ratio, convex hull ratio, spatial presentation format, and temporal presentation format each significantly moderated associations between numerosity ratio and accuracy when included independently and as such were combined into a single model. Results are shown in the second column of **Table 3**. Significant associations remained for surface area ratio, spatial presentation format, and temporal presentation format.

Numerosity ratio effects were significantly larger on trials where surface area ratio and numerosity ratio were less congruent (see **Figure 2**). In other words, the congruency between surface area and numerosity was most strongly related to accuracy on more difficult trials (i.e., trials with smaller numerosity ratio) and was not significantly related to performance on the easiest trials (i.e., trials with larger numerosity ratios).

Additionally, numerosity ratio effects were significantly larger on spatially separated compared to overlapping trials (see **Figure 3**). The differences in accuracy between spatially separated



and overlapping trials favoring separated trials were largest for easier trials (i.e., trials with larger numerosity ratios) compared to more difficult trials (i.e., trials with smaller numerosity ratios).

Finally, numerosity ratio effects were significantly larger on sequentially compared to simultaneously presented trials (see

Figure 4). The difference in odds ratios among sequentially and simultaneously presented trials favoring sequential trials were largest among easier trials (i.e., trials with larger numerosity ratios) and were actually non-significant on the most difficult trials (i.e., trials with smaller numerosity ratios).

Math Fluency Interactions

Math fluency scores were then included as a predictor of the level-1 coefficient on numerosity ratio to represent a cross-level interaction between numerosity ratio and math ability. Model estimates are shown in the third column of **Table 3**. In addition to the positive main effects of math fluency on overall accuracy (i.e., intercepts), math fluency significantly predicted individuals' numerosity ratio slopes, such that for participants with higher math fluency scores, associations between numerosity ratio and accuracy were higher (see **Figure 5**). Participants with higher math scores appear more responsive to number than participants with lower math scores. In other words, math fluency was more positively related to performance on easier trials (i.e., trials with larger numerosity ratios) but was not significantly related to performance on harder trials (i.e., trials with smaller numerosity ratios).

Average Dot Size and Density as Trial-Level Predictors

Results from these models using measures of average dot size ratio and density ratio as predictors of accuracy are shown in **Table 4**. Consistent with the results described above, numerosity ratio remained a significant predictor of individuals' performance across all model specifications. However, it is notable that both average dot size and display density were significant predictors of performance as well, as participants were more accurate on trials in which dot size congruency was higher and density congruency was lower. Dot size also significantly moderated numerosity ratio effects, such that numerical ratio effects were smaller on trials in which dot size was more congruent, consistent with

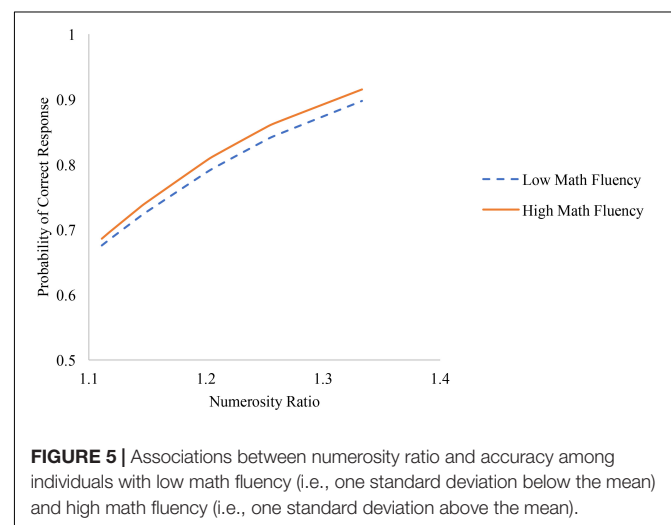


TABLE 4 | Results of alternative two-level logistic hierarchical linear models predicting trial-level accuracy on the non-symbolic number comparison task (1 = correct response) from trial-level characteristics, including average dot size and density, and participant-level characteristics.

Fixed Effects	OR (SE)	OR (SE)	OR (SE)
Trial-Level Characteristics			
Numerosity Ratio	1.90*** (0.03)	2.06*** (0.07)	2.06*** (0.07)
Dot Size Ratio	1.52*** (0.11)	2.89*** (0.68)	2.89*** (0.68)
Density Ratio	0.43*** (0.04)	0.43** (0.11)	0.43** (0.11)
Spatially Overlapping Presentation	0.66*** (0.02)	1.21** (0.08)	1.21** (0.08)
Sequential Presentation	1.12*** (0.02)	0.96 (0.06)	0.95 (0.06)
Participant-Level Characteristics			
Math Fluency	1.004* (0.002)	1.004* (0.002)	0.999 (0.003)
Age	0.99 (0.01)	0.99 (0.01)	0.99 (0.01)
Female	0.97 (0.05)	0.97 (0.05)	0.97 (0.05)
Visuospatial Short-Term Memory	1.003 (0.04)	1.003 (0.04)	1.004 (0.04)
Phonological Working Memory	1.02 (0.02)	1.02 (0.02)	1.02 (0.02)
Trial-Level Interactions			
Dot Size Ratio * Numerosity Ratio		0.77** (0.07)	0.77** (0.07)
Density Ratio * Numerosity Ratio		1.03 (0.12)	1.03 (0.13)
Spatially Overlapping Presentation * Numerosity Ratio		0.73*** (0.02)	0.73*** (0.02)
Sequential Presentation * Numerosity Ratio		1.09** (0.03)	1.09** (0.03)
Cross-Level Interactions			
Math Fluency * Numerosity Ratio			1.003* (0.001)
Intercept	1.20*** (0.06)	1.04 (0.08)	1.03 (0.08)
Random effect			
	Estimate (SE)	Estimate (SE)	Estimate (SE)
Intercept	0.06 (0.01)	0.06 (0.01)	0.06 (0.01)

Values shown in the table are odds ratios and their standard errors. Numerosity ratio was centered at 1, and dot size ratio and density ratio were natural log transformed. Math fluency scores were mean-centered prior to estimating models. † $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

the cumulative surface area interaction shown in **Figure 2**. Importantly, the inclusion of these alternative metrics of visual confounds in the stimuli did not change the remainder of the findings, including numerical ratio interactions with spatial or temporal presentation format or math fluency scores.

DISCUSSION

Issues surrounding (1) the measurement of the ANS and (2) the relation between individual differences in ANS acuity and math performance are both highly debated (Gebuis et al., 2016; Leibovich and Ansari, 2016; Leibovich et al., 2016). To our knowledge, we are the first to utilize hierarchical linear models (HLMs) to study the ANS and to simultaneously examine differences in non-symbolic number comparison performance from person to person and from trial to trial. This approach allowed us to account for the nested structure of our data, to account for variance in trial-level and participant-level variables at the same time, and to learn the distribution of effects across people by modeling the participant-level characteristics as random effects rather than fixed effects. Below we discuss our findings regarding the role of numerosity ratio, perceptual continuous dimensions, presentation format, and participants' math ability on non-symbolic number comparison trial-level accuracy, and

the role of these variables in modulating numerosity ratio effects.

Effects of Numerosity Ratio

Replicating numerous studies (Dehaene, 1992; Cantlon and Brannon, 2006; Libertus et al., 2007; Halberda and Feigenson, 2008; Halberda et al., 2008; Soltész et al., 2010; Inglis et al., 2011; Dewind and Brannon, 2012; Price et al., 2012; Agrillo et al., 2013), we found that participants were more accurate on trials with easier numerosity ratios compared to more difficult numerosity ratios, i.e., they were more likely to correctly identify the larger quantity as the relative difference between the two numerosities became larger. Importantly, numerosity ratio was a highly significant predictor of accuracy above and beyond all measured trial-level variables, including convex hull ratio, surface area ratio, average dot size ratio, density ratio and variations in spatial and temporal presentation format of the stimuli. Thus, our finding is in line with prior work that suggests number, or numerosity ratio, is a highly salient dimension of non-symbolic stimuli (Cordes and Brannon, 2009; Libertus et al., 2014; DeWind et al., 2015; Park et al., 2016; Starr et al., 2017). Our study also extends this work by additionally controlling for participant-level variables, including participants' age, gender, visuospatial short-term memory, phonological working memory, and math ability. Numerosity ratio remained a highly significant predictor of accuracy above and beyond all measured participant-level

variables. These findings are particularly noteworthy given recent evidence indicating that critical non-numerical cues such as convex hull are not controlled for in the stimulus design of Panamath (Clayton et al., 2015). Importantly, numerosity ratio also remained a significant predictor of accuracy on all trial types (although not equally so, as will be discussed below), demonstrating that across task specifications, numerical information is related to performance. Thus, numerosity ratio, the critical marker of the ANS, seems to be an independent and robust indicator of non-symbolic number comparison performance.

Effects of Continuous Dimensions on Non-symbolic Number Comparison

Our results indicate that our participants' accuracy on the non-symbolic number comparison task cannot be explained entirely by numerosity ratio; certain trial-level characteristic of the dot arrays contribute to peoples' ability to compare numerosities. On the one hand, the cumulative surface area of the dot arrays (or alternatively the average individual size of a dot in the arrays) was significantly associated with accuracy on the non-symbolic number comparison task, controlling for numerosity ratio and all other trial-level and participant-level characteristics. Specifically, increasing surface area congruency (the array with the larger number is also the array with the larger cumulative surface area), increased participants' odds of responding correctly.

Cumulative surface area ratio and average individual dot size ratio also moderated the association between numerosity ratio and accuracy. On trials with easier numerosity ratios, participants performed similarly regardless of whether there was high surface area/dot size congruency or low surface area/dot size congruency, but on trials with more difficult numerosity ratios, participants were more accurate when there was high surface area/dot size congruency. While participants may be able to indicate the larger numerosity on easy trials by simply relying on numerosity as their primary cue, they may rely on other cues, namely surface area or dot size, to a greater extent as the numerosity ratio becomes more difficult to discriminate. When the numerosity ratio of the trial is difficult, using surface area or dot size provides a potentially useful, although not perfect, indicator that there are more items in the array, and leads to more accurate performance when the surface area or dot size information has high congruency with the numerosity information. This explanation is in line with the Signal Clarity Hypothesis, which states that the clarity of numerosity estimates can be supported by dimensions of continuous quantity when they co-vary with or are redundant with number (Cantrell and Smith, 2013; Cantrell et al., 2015). These findings are consistent with past work demonstrating that participants tend to be more accurate on surface area congruent trials compared to incongruent trials (e.g., Dewind and Brannon, 2012) but also extend this work by addressing how and when these congruency effects are likely to come into play.

On the other hand, increases in convex hull and density congruency also significantly predicted increases in accuracy. Participants were overall more accurate when the array with

the larger number also had the larger convex hull or was denser, holding numerosity ratio and all other trial-level and participant-level characteristics constant. Increases in convex hull congruency were even more predictive of accuracy than increases in surface area congruency (133 and 52% increase in the odds of responding correctly, respectively). This result supports previous studies that describe the influence of convex hull on non-symbolic number comparison performance (Clayton et al., 2015; DeWind and Brannon, 2016) and those demonstrating that, for adults, convex hull may be a more salient dimension than surface area on these tasks (Gilmore et al., 2016). In contrast, density ratio was less predictive of accuracy than average dot size ratio possibly because extracting information about individual dot size may be easier than extracting information about cumulative surface area (Cordes and Brannon, 2008).

Effects of Spatial and Temporal Variations in Stimulus Presentation Format

We also found a significant influence of both spatial separation and the temporal aspects of the stimulus presentation on participants' accuracy. First, participants were more accurate on trials when the arrays were presented with spatial separation (52% higher odds of responding correctly) compared to spatial overlap, mirroring previous findings in the literature (Price et al., 2012; Norris and Castronovo, 2016). Together, these studies suggest that spatially overlapping displays are more difficult to compare, most likely because they require additional cognitive processing to visually segment the two arrays. Our study also provides new evidence that the spacing of the presentation format (separated or overlapping) moderates the association between numerosity ratio and participants' accuracy, such that the benefit of spatially separated compared to spatially overlapping displays is greater on trials with easier numerosity ratios. One possible explanation for this result is that participants use different strategies when performing number comparisons of spatially separated and spatially overlapping arrays and that the use of these strategies is affected by numerosity ratio. However, future studies are needed to directly test this hypothesis.

Additionally, participants in our sample were significantly more accurate on trials when the arrays were presented sequentially compared to simultaneously. The benefit of sequential trials found here is opposite of the finding by Smets et al. (2016) who reported an advantage for simultaneously presented trials. It is possible that performance differences across the two studies are driven by presentation time differences; in both studies, each array in the sequentially presented trials was displayed for 750 ms, but Smets and colleagues had the arguably more difficult task because they included a 500-ms delay between the two arrays. It should also be noted that in our sample, the benefit of sequential trials over simultaneous trials was a relatively small effect (12% higher odds of responding correctly). Future studies manipulating this delay time would be instrumental in unpacking these findings and exploring how non-symbolic representations are maintained.

Mirroring the interaction we found for variations in spatial stimulus presentation with numerosity ratio, we found an interaction between temporal variations in stimulus presentation format and numerosity ratio. Participants showed greater benefit of sequential compared to simultaneous presentation on trials with easier numerosity ratios. Again, one possible explanation for this result may be that participants use different strategies when performing number comparisons of sequentially and simultaneously presented arrays and that the use of these strategies is affected by numerosity ratio. One possible approach to test this hypothesis would be to use eye tracking to compare participants' scanning patterns as they process the same arrays in the two conditions (see Pailian and Halberda, 2015, for a similar approach to compare differences between number and area comparisons). Another possible explanation is that the sequential presentation enables participants to form a solid representation of the first numerosity before comparing it to the second. However, this representational strength is more beneficial in an easy ratio when there is little overlap between the two representations of the numerosities.

Effects of Participant-Level Characteristics

In addition to examining trial-level predictors of accuracy on this non-symbolic number comparison task, we were also interested in identifying participant-level predictors of individuals' accuracy in this task. Consistent with past research (e.g., Halberda and Feigenson, 2008; Halberda et al., 2008, 2012; Inglis et al., 2011; Libertus et al., 2011, 2012; Mazzocco et al., 2011; Dewind and Brannon, 2012; Lourenco et al., 2012; Bonny and Lourenco, 2013; Guillaume et al., 2013; Keller and Libertus, 2015; Braham and Libertus, 2017, 2018), we found that participants with higher math fluency scores tended to have higher ANS acuity, as indicated by higher average odds of responding correctly. This association was quite small in magnitude (a standard deviation increase in fluency predicted a 7% increase in odds of correctly responding, which is equivalent to the difference between 60 and 62% probability) but was seen when controlling for domain-general cognitive skills.

Due to model specifications, math scores were included as a predictor of ANS performance rather than ANS acuity predicting math, as is typically seen in the literature (e.g., Gilmore et al., 2010; Libertus et al., 2011, 2013a; Mazzocco et al., 2011; Starr et al., 2013; Keller and Libertus, 2015). However, growing evidence indicates that these associations between math skills and the ANS may be bidirectional, such that math skills may actually support the development of the ANS. Piazza et al. (2013) demonstrated that adult speakers of Mundurukú, a language that lacks number words beyond five and therefore severely limits the mathematical concepts that speakers can articulate, have less precise representations of approximate quantities than do individuals from Western cultures who speak languages that include number words. Similarly, evidence with Western adults suggests that formal math education is associated with greater precision of the ANS (Nys et al., 2013; Lindskog et al., 2014). Furthermore, two recent studies utilized cross-lagged

longitudinal designs have shown that children's math skills predict later ANS acuity, even when controlling for earlier ANS acuity, suggesting that math may relate to changes in the ANS over time (authors, under review; Mussolin et al., 2014; but see He et al., 2016). As such, associations between the ANS and math may in fact be bidirectional, at least in early childhood. However, the present study was cross-sectional in nature, and so our findings cannot inform these hypotheses. Instead, our seemingly directional pathways simply reflect patterns of correlations across individuals.

Finally, we found that ratio effects on accuracy were moderated by math ability, such that individuals with higher math fluency were more responsive to ratio. These results indicate that individuals with stronger math skills may be more influenced by numerical information provided in the stimuli, although math ability did not significantly moderate associations between non-numerical information and accuracy, indicating that participants with stronger math skills did not necessarily rely on numerical information more and non-numerical information less. As such, more research is needed to unpack the ways that adults with varying levels of math skills process these displays and discriminate between quantities.

Limitations and Conclusions

There are several limitations of this study that should be addressed in future research. First, unlike the methods of DeWind et al. (2015), we did not systematically vary surface area/dot size and convex hull/density ratios to have equivalent ranges. Thus, we acknowledge that our findings about the relative salience of numerosity ratio, cumulative surface area ratio, average dot size ratio, convex hull ratio, and density ratio, are constrained by the range of variability of these ratios in our stimuli. An important avenue for future research will be to combine the stimuli of Dewind and colleagues with our HLM analyses, which account for both trial-level and participant-level characteristics simultaneously. Second, our measure of participants' math ability was limited to an assessment of speeded mental arithmetic. In light of work suggesting that ANS acuity may be differentially related to various aspects of math, and specifically that mental arithmetic may be more strongly related to ANS acuity than written arithmetic (Schneider et al., 2016; Braham and Libertus, 2018), future research is needed to follow up on this analysis approach using varied and more broad measures of math ability.

To summarize, our results indicate that certain trial-level confounds of the dot arrays, including cumulative surface area, average individual dot size, convex hull and density as well as spatial and temporal variations of the stimulus presentation, and certain characteristics of the participants, namely math ability, contribute to the ability to compare numerosities on the non-symbolic comparison task. Yet numerosity ratio, the critical marker of the ANS, remained a highly significant predictor of accuracy even when all other trial-level and participant-level characteristics were included in our models. Thus, our findings add further support for the argument that, although some trial-level confounds affect number judgments, numerosity ratio seems to be an independent and critical feature of non-symbolic number comparison performance, even across individuals with

varying levels of math ability and domain-general cognitive skills.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the University of Pittsburgh's Institutional Review Board. The protocol was approved by the University of Pittsburgh's Institutional Review Board (PRO13090407).

REFERENCES

- Agrillo, C., Piffer, L., and Adriano, A. (2013). Individual differences in non-symbolic numerical abilities predict mathematical achievements but contradict ATOM. *Behav. Brain Funct.* 9, 1–14. doi: 10.1186/1744-9081-9-26
- Allik, J., and Tuulmets, T. (1991). Occupancy model of perceived numerosity. *Percept. Psychophys.* 49, 303–314. doi: 10.3758/BF03205986
- Ansari, D., Lyons, I. M., van Eimeren, L., and Xu, F. (2007). Linking visual attention and number processing in the brain: the role of the temporo-parietal junction in small and large symbolic and nonsymbolic number comparison. *J. Cogn. Neurosci.* 19, 1845–1853. doi: 10.1162/jocn.2007.19.11.1845
- Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., and Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition* 98, 199–222. doi: 10.1016/j.cognition.2004.09.011
- Bonny, J. W. J., and Lourenco, S. F. S. (2013). The approximate number system and its relation to early math achievement: evidence from the preschool years. *J. Exp. Child Psychol.* 114, 375–388. doi: 10.1016/j.jecp.2012.09.015
- Braham, E. J., and Libertus, M. E. (2017). Intergenerational associations in numerical approximation and mathematical abilities. *Dev. Sci.* 20:5. doi: 10.1111/desc.12436
- Braham, E. J., and Libertus, M. E. (2018). When approximate number acuity predicts math performance: the moderating role of math anxiety. *PLoS One* 13:e0195696. doi: 10.1371/journal.pone.0195696
- Brankaer, C., Ghesquière, P., and De Smedt, B. (2014). Children's mapping between non-symbolic and symbolic numerical magnitudes and its association with timed and untimed tests of mathematics achievement. *PLoS One* 9:e93565. doi: 10.1371/journal.pone.0093565
- Brown, B. R., and Rebbin, T. J. (1970). Simultaneous vs sequential discriminations of Markov-generated stimuli. *Percept. Psychophys.* 8, 353–357. doi: 10.3758/BF03212607
- Bugden, S., and Ansari, D. (2016). Probing the nature of deficits in the "Approximate Number System" in children with persistent Developmental Dyscalculia. *Dev. Sci.* 19, 817–833. doi: 10.1111/desc.12324
- Cantlon, J. F., and Brannon, E. M. (2006). Shared system for ordering small and large numbers in and humans monkeys. *Psychol. Sci.* 17, 401–406. doi: 10.1111/j.1467-9280.2006.01719.x
- Cantrell, L., Boyer, T. W., Cordes, S., and Smith, L. B. (2015). Signal clarity: an account of the variability in infant quantity discrimination tasks. *Dev. Sci.* 18, 877–893. doi: 10.1111/desc.12283
- Cantrell, L., and Smith, L. B. (2013). Open questions and a proposal: a critical review of the evidence on infant numerical abilities. *Cognition* 128, 331–352. doi: 10.1016/j.cognition.2013.04.008
- Castronovo, J., and Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS One* 7:e33832. doi: 10.1371/journal.pone.0033832
- Chen, Q., and Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychologica* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Clayton, S., Gilmore, C., and Inglis, M. (2015). Dot comparison stimuli are not all alike: the effect of different visual controls on ANS measurement. *Acta Psychol.* 161, 177–184. doi: 10.1016/j.actpsy.2015.09.007
- Cordes, S., and Brannon, E. M. (2008). Quantitative competencies in infancy. *Dev. Sci.* 11, 803–808. doi: 10.1111/j.1467-7687.2008.00770.x

All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

EB and ML conceptualized the study. EB collected the data. LE and EB analyzed the data and drafted the manuscript. All authors contributed to editing, reviewing, and approving the final manuscript.

- Cordes, S., and Brannon, E. M. (2009). The relative salience of discrete and continuous quantity in young infants. *Dev. Sci.* 12, 453–463. doi: 10.1111/j.1467-7687.2008.00781.x.The
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N
- DeWind, N. K., Adams, G. K., Platt, M. L., and Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition* 142, 247–265. doi: 10.1016/j.cognition.2015.05.016
- Dewind, N. K., and Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Front. Hum. Neurosci.* 6:68. doi: 10.3389/fnhum.2012.00068
- DeWind, N. K., and Brannon, E. M. (2016). Significant inter-test reliability across approximate number system assessments. *Front. Psychol.* 7:310. doi: 10.3389/fpsyg.2016.00310
- Dillon, M. R., Kannan, H., Dean, J. T., Spelke, E. S., and Duflo, E. (2017). Cognitive science in the field: a preschool intervention durably enhances Intuitive But Not Formal Mathematics. *Science* 55, 47–55. doi: 10.1126/science.aal4724
- Durgin, F. H. (1995). Texture density adaptation and the perceived numerosity and distribution of texture. *J. Exp. Psychol.* 21, 149–169. doi: 10.1037/0096-1523.21.1.149
- Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53–72. doi: 10.1016/j.jecp.2014.01.013
- Feigenson, L., Libertus, M. E., and Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Dev. Perspect.* 7, 74–79. doi: 10.1111/cdep.12019
- Frick, R. W. (1985). Testing visual short-term memory: simultaneous versus sequential presentations. *Mem. Cogn.* 13, 346–356. doi: 10.3758/BF03202502
- Fuhs, M. W., and McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: contributions of inhibitory control. *Dev. Sci.* 16, 136–148. doi: 10.1111/desc.12013
- Geary, D. C., vanMarle, K., Chu, F. W., Rouder, J., Hoard, M. K., and Nugent, L. (2018). Early conceptual understanding of cardinality predicts superior school-entry number-system knowledge. *Psychol. Sci.* 29, 191–205. doi: 10.1177/0956797617729817
- Gebuis, T., Cohen Kadosh, R., and Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: a critical review. *Acta Psychol.* 171, 17–35. doi: 10.1016/j.actpsy.2016.09.003
- Gebuis, T., and Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *J. Exp. Psychol. Gen.* 141, 642–648. doi: 10.1037/a0026218
- Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., et al. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS One* 8:e67374. doi: 10.1371/journal.pone.0067374
- Gilmore, C., Cragg, L., Hogan, G., and Inglis, M. (2016). Congruency effects in dot comparison tasks: convex hull is more important than dot area. *J. Cogn. Psychol.* 28, 923–931. doi: 10.1080/20445911.2016.1221828
- Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition* 115, 394–406. doi: 10.1016/j.cognition.2010.02.002

- Guillaume, M., Nys, J., Mussolin, C., and Content, A. (2013). Differences in the acuity of the approximate number system in adults: the effect of mathematical ability. *Acta Psychol.* 144, 506–512. doi: 10.1016/j.actpsy.2013.09.001
- Haist, F., Wazny, J. H., Toomarian, E., and Adamo, M. (2015). Development of brain systems for nonsymbolic numerosity and the relationship to formal math academic achievement. *Hum. Brain Mapp.* 36, 804–826. doi: 10.1002/hbm.22666
- Halberda, J., and Feigenson, L. (2008). Developmental change in the acuity of the “Number Sense”: the Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Dev. Psychol.* 44, 1457–1465. doi: 10.1037/a0012682
- Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11116–11120. doi: 10.1073/pnas.1200196109
- Halberda, J., Mazzocco, M. M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with Maths achievement. *Nature* 455, 665–668. doi: 10.1038/nature07246
- Hayashi, M. J., Kanai, R., Tanabe, H. C., Yoshida, Y., Carlson, S., Walsh, V., et al. (2013). Interaction of numerosity and time in prefrontal and parietal cortex. *J. Neurosci.* 33, 883–893. doi: 10.1523/JNEUROSCI.6257-11.2013
- He, Y., Zhou, X., Shi, D., Song, H., Zhang, H., and Shi, J. (2016). New evidence on causal relationship between approximate number system (ANS) acuity and arithmetic ability in elementary-school students: a longitudinal cross-lagged analysis. *Front. Psychol.* 7:1052. doi: 10.3389/fpsyg.2016.01052
- Henik, A., Gliksmann, Y., Kallai, A., and Leibovich, T. (2017). Size perception and the foundation of numerical processing. *Curr. Dir. Psychol. Sci.* 26, 45–51. doi: 10.1177/09637214166671323
- Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children’s mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001
- Hyde, D. C., Khanum, S., and Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition* 131, 92–107. doi: 10.1016/j.cognition.2013.12.007
- Inglis, M., Attridge, N., Batchelor, S., and Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: but only in children. *Psychon. Bull. Rev.* 18, 1222–1229. doi: 10.3758/s13423-011-0154-1
- Inglis, M., and Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychol.* 145, 147–155. doi: 10.1016/j.actpsy.2013.11.009
- Keller, L., and Libertus, M. (2015). Inhibitory control may not explain the link between approximation and math abilities in kindergarteners from middle class families. *Front. Psychol.* 6:685. doi: 10.3389/fpsyg.2015.00685
- Kolkman, M. E., Kroesbergen, E. H., and Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learn. Instr.* 25, 95–103. doi: 10.1016/j.learninstruc.2012.12.001
- Leibovich, T., and Ansari, D. (2016). The symbol-grounding problem in numerical cognition: a review of theory, evidence, and outstanding questions. *Can. J. Exp. Psychol.* 70, 12–23. doi: 10.1037/cep0000070
- Leibovich, T., Katzin, N., Harel, M., and Henik, A. (2016). From “sense of number” to “sense of magnitude” – The role of continuous magnitudes in numerical cognition. *Behav. Brain Sci.* 40:164. doi: 10.1017/S0140525X16000960
- Libertus, M. E., Feigenson, L., and Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Dev. Sci.* 14, 1292–1300. doi: 10.1111/j.1467-7687.2011.01080.x
- Libertus, M. E., Feigenson, L., and Halberda, J. (2013a). Is approximate number precision a stable predictor of math ability? *Learn. Individ. Diff.* 25, 126–133. doi: 10.1016/j.lindif.2013.02.001
- Libertus, M. E., Feigenson, L., and Halberda, J. (2013b). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *J. Exp. Child Psychol.* 116, 829–838. doi: 10.1016/j.jecp.2013.08.003
- Libertus, M. E., Odic, D., and Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychol.* 141, 373–379. doi: 10.1016/j.actpsy.2012.09.009
- Libertus, M. E., Starr, A., and Brannon, E. M. (2014). Number trumps area for 7-month-old infants. *Dev. Psychol.* 50, 108–112. doi: 10.1037/a0032986
- Libertus, M. E., Woldorff, M. G., and Brannon, E. M. (2007). Electrophysiological evidence for notation independence in numerical processing. *Behav. Brain Funct.* 3:1. doi: 10.1186/1744-9081-3-1
- Lindskog, M., Winman, A., and Juslin, P. (2014). The association between higher education and approximate number system acuity. *Front. Psychol.* 5:462. doi: 10.3389/fpsyg.2014.00462
- Lindskog, M., Winman, A., Juslin, P., and Poom, L. (2013). Measuring acuity of the approximate number system reliably and validly: the evaluation of an adaptive test procedure. *Front. Psychol.* 4:510. doi: 10.3389/fpsyg.2013.00510
- Lindskog, M., Winman, A., and Poom, L. (2017). Individual differences in nonverbal number skills predict math anxiety. *Cognition* 159, 156–162. doi: 10.1016/j.cognition.2016.11.014
- Lourenco, S. F., Bonny, J. W., Fernandez, E. P., and Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proc. Natl. Acad. Sci.* 109, 18737–18742. doi: 10.1073/pnas.1207212109
- Lukowski, S. L., Rosenberg-Lee, M., Thompson, L. A., Hart, S. A., Willcutt, E. G., Olson, R. K., et al. (2017). Approximate number sense shares etiological overlap with mathematics and general cognitive ability. *Intelligence* 65, 67–74. doi: 10.1016/j.intell.2017.08.005
- Lyons, I. M., and Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition* 121, 256–261. doi: 10.1016/j.cognition.2011.07.009
- Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011). Preschoolers’ precision of the approximate number system predicts later school mathematics performance. *PLoS One* 6:e23749. doi: 10.1371/journal.pone.0023749
- Mussolin, C., Nys, J., Content, A., and Leybaert, J. (2014). Symbolic number abilities predict later approximate number system acuity in preschool children. *PLoS One* 9:e91839. doi: 10.1371/journal.pone.0091839
- Norris, J. E., and Castronovo, J. (2016). Dot display affects approximate number system acuity and relationships with mathematical achievement and inhibitory control. *PLoS One* 11:e0155543. doi: 10.1371/journal.pone.0155543
- Norris, J. E., McGeown, W. J., Guerrini, C., and Castronovo, J. (2015). Aging and the number sense: preserved basic non-symbolic numerical processing and enhanced basic symbolic processing. *Front. Psychol.* 6:999. doi: 10.3389/fpsyg.2015.00999
- Nys, J., Ventura, P., Fernandes, T., Querido, L., Leybaert, J., and Content, A. (2013). Does math education modify the approximate number system? A comparison of schooled and unschooled adults. *Trends Neurosci. Educ.* 2, 13–22. doi: 10.1111/bjdp.12234
- Odic, D., and Halberda, J. (2015). Eye movements reveal distinct encoding patterns for number and cumulative surface area in random dot arrays. *J. vis.* 15:5. doi: 10.1167/15.15.5
- Pailian, H., and Halberda, J. (2015). The reliability and internal consistency of one-shot and flicker change detection for measuring individual differences in visual working memory capacity. *Mem. Cognit.* 43, 397–420. doi: 10.3758/s13421-014-0492-0
- Park, J., and Brannon, E. M. (2014). Improving arithmetic performance with number sense training: an investigation of underlying mechanism. *Cognition* 133, 188–200. doi: 10.1016/j.cognition.2014.06.011
- Park, J., Dewind, N. K., Woldorff, M. G., and Brannon, E. M. (2016). Rapid and direct encoding of numerosity in the visual stream. *Cereb. Cortex* 26, 748–763. doi: 10.1093/cercor/bhv017
- Patalano, A. L., Saltiel, J. R., Machlin, L., and Barth, H. (2015). The role of numeracy and approximate number system acuity in predicting value and probability distortion. *Psychonom. Bull. Rev.* 22, 1820–1829. doi: 10.3758/s13423-015-0849-9
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012
- Piazza, M., Pica, P., Izard, V., Spelke, E. S., and Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychol. Sci.* 24, 1037–1043. doi: 10.1177/0956797612464057
- Pinheiro-Chagas, P., Krinzinger, H., Lonnemann, J., Wood, G., Knops, A., Krinzinger, H., et al. (2014). In how many ways is the approximate number system associated with exact calculation? *PLoS One* 9:e111155. doi: 10.1371/journal.pone.0111155
- Price, G. R., Palmer, D., Battista, C., and Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: reliability and validity of different task variants and

- outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychol.* 140, 50–57. doi: 10.1016/j.actpsy.2012.02.008
- Purpura, D. J., and Logan, J. A. R. (2015). The nonlinear relations of the approximate number system and mathematical language to early mathematics development. *Dev. Psychol.* 51, 1717–1724. doi: 10.1037/dev0000055
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: what underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418–431. doi: 10.1016/j.jecp.2012.10.012
- Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S. S., Stricker, J., et al. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:3. doi: 10.1111/desc.12372
- Smets, K., Gebuis, T., Defever, E., and Reynvoet, B. (2014). Concurrent validity of approximate number sense tasks in adults and children. *Acta Psychol.* 150, 120–128. doi: 10.1016/j.actpsy.2014.05.001
- Smets, K., Moors, P., and Reynvoet, B. (2016). Effects of presentation type and visual control in numerosity discrimination: implications for number processing? *Front. Psychol.* 7:66. doi: 10.3389/fpsyg.2016.00066
- Soltész, F., Szucs, D., and Szucs, L. (2010). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: a developmental study. *Behav. Brain Funct.* 6, 1–14. doi: 10.1186/1744-9081-6-13
- Starr, A., DeWind, N. K., and Brannon, E. M. (2017). The contributions of the number sense and symbolic math achievement. *Cognition* 168, 222–233. doi: 10.1016/j.cognition.2017.07.004
- Starr, A., Libertus, M. E., and Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proc. Natl. Acad. Sci. U.S.A.* 110, 1–5. doi: 10.1073/pnas.1302751110
- Tokita, M., and Ishiguchi, A. (2010). How might the discrepancy in the effects of perceptual variables on numerosity judgment be reconciled? *Attent. Percept. Psychophys.* 72, 1839–1853. doi: 10.3758/APP.72.7.1839
- van Marle, K., Chu, F. W., Li, Y., and Geary, D. C. (2014). Acuity of the approximate number system and preschoolers' quantitative development. *Dev. Sci.* 17, 492–505. doi: 10.1111/desc.12143
- Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*, 3rd Edn. Itasca, IL: Riverside Publishing Company.
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Braham, Elliott and Libertus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Comparing Numerical Comparison Tasks: A Meta-Analysis of the Variability of the Weber Fraction Relative to the Generation Algorithm

Mathieu Guillaume^{1*} and Amandine Van Rinsveld²

¹ Cognitive Science and Assessment Institute (COISA), University of Luxembourg, Luxembourg, Luxembourg, ² Centre for Research in Cognitive Neuroscience (CRCN), Université Libre de Bruxelles, Brussels, Belgium

OPEN ACCESS

Edited by:

Jingguang Li,
Dali University, China

Reviewed by:

Wei Wei,
Shanghai Normal University, China
Zonglei Zhen,
Beijing Normal University, China

*Correspondence:

Mathieu Guillaume
mathieu.guillaume@uni.lu

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 15 April 2018

Accepted: 22 August 2018

Published: 11 September 2018

Citation:

Guillaume M and Van Rinsveld A
(2018) Comparing Numerical
Comparison Tasks: A Meta-Analysis of
the Variability of the Weber Fraction
Relative to the Generation Algorithm.
Front. Psychol. 9:1694.
doi: 10.3389/fpsyg.2018.01694

Since more than 15 years, researchers have been expressing their interest in evaluating the Approximate Number System (ANS) and its potential influence on cognitive skills involving number processing, such as arithmetic. Although many studies reported significant and predictive relations between ANS and arithmetic abilities, there has recently been an increasing amount of published data that failed to replicate such relationship. Inconsistencies lead many researchers to question the validity of the assessment of the ANS itself. In the current meta-analysis of over 68 experimental studies published between 2004 and 2017, we show that the mean value of the Weber fraction (w), the minimal amount of change in magnitude to detect a difference, is very heterogeneous across the literature. Within young adults, w might range from <10 to more than 60, which is critical for its validity for research and diagnostic purposes. We illustrate here the concern that different methods controlling for non-numerical dimensions lead to substantially variable performance. Nevertheless, studies that referred to the exact same method (e.g., Panamath) showed high consistency among them, which is reassuring. We are thus encouraging researchers only to compare what is comparable and to avoid considering the Weber fraction as an abstract parameter independent from the context. Eventually, we observed that all reported correlation coefficients between the value of w and general accuracy were very high. Such result calls into question the relevance of computing and reporting at all the Weber fraction. We are thus in disfavor of the systematic use of the Weber fraction, to discourage any temptation to compare given data to some values of w reported from different tasks and generation algorithms.

Keywords: Approximate Number System, number sense, meta-analysis, methodology, Weber fraction

Over 20 years ago, Dehaene made the hypothesis that Humans possess a *Number Sense*, a biologically determined ability that allows us to represent and manipulate large numerical quantities (Dehaene, 1997). This numerical intuition is largely considered as relying on a cognitive system specifically dedicated to number processing called the *Approximate Number System* (ANS, Feigenson et al., 2004; see also Núñez, 2017; for an interesting terminological criticism). The crucial property of such cognitive system is the scalar variability of numerical approximations: numerical estimates of larger quantities are indeed more variable (Platt and Johnson, 1971; Gallistel and Gelman, 2000). Accordingly, the acuity of numerical discriminative processes handling two amounts is not absolute, but relative to the numerical ratio between the considered quantities (i.e., distinguishing 10 from 20 elements is easier than distinguishing 110 from 120 items). Mental number representations were thus hypothesized to go through a logarithmic compression following the Weber-Fechner law (Dehaene, 2003; but see Cantlon et al., 2009; Cicchini et al., 2014; and Piantadosi, 2016).

In order to assess these logarithmic representations, Piazza et al. (2004) were among the first to characterize performance (as well as brain activity) in a numerical discrimination task with the help of a measure directly related to the Weber-Fechner law, the Weber fraction. The Weber fraction is the ratio between the amount just noticeably different from a given magnitude, and the magnitude itself (w , see Stevens, 1957; Van Oeffelen and Vos, 1982). From a psychophysical perspective, the Weber fraction can be defined as the noise constant-proportionality parameter fitting the discrimination behavior during a numerical comparison task (see Barth et al., 2006, Appendix B). As a constant scaling ratio, the Weber fraction has the advantage of explicitly depicting the scalar variability across mental representations, which might fluctuate between individuals (see Whalen et al., 1999). More critically to the purpose of the current meta-analysis, this w parameter was heavily popularized in the literature as a direct measure of specific numerical quantity processes by some influential studies (e.g., Pica et al., 2004; Piazza et al., 2004). Subsequently, w was widely investigated as an individual property that is only subject to significant developmental changes across the lifespan (Halberda et al., 2012) and to refinement through formal instruction (Piazza et al., 2013). For a given age within a given population, w was thus considered as a stable predictor of more complex numerical processing such as math ability (Halberda et al., 2008), as well as a crucial clinical predictor of Mathematical Learning Disability (e.g., Mazzocco et al., 2011).

However, some authors recently questioned the stability of the Weber fraction. Due to the substantial amount of studies that were conducted following Halberda et al. study (2008), there were indeed many reports of failure in observing significant relationship between w and math ability (e.g., Price et al., 2012; Gilmore et al., 2013; Sasanguie et al., 2013). This raised some theoretical concerns (e.g., Gebuis et al., 2016; Leibovich et al., 2016; Núñez, 2017), as well as many methodological issues (see Dietrich et al., 2015b; for a review). Among these issues, many studies showed that the assessment of ANS acuity, and the measure of w itself, are not independent of interference from low level visual cues that are intrinsically confounded with

numerical quantities, and they revealed that w is not consistent nor reliable across different tasks (Clayton et al., 2015, 2018; Bugden and Ansari, 2016; Guillaume et al., 2016). Some authors subsequently argued that the procedure used to generate visual arrays substantially influence participant behavior, and therefore the evaluation of w (Inglis and Gilmore, 2014; Clayton et al., 2015; Smets et al., 2015, 2016). In other words, the Weber fraction does not seem to be a stable psychophysiological parameter devoid of context; w can in fact be variable within one subject as a function of the task and the stimulus properties (but see Julio-Costa et al., 2015; DeWind and Brannon, 2016; for contradicting evidence).

Inglis and Gilmore (2014) went further by experimentally assessing the validity and the reliability of the Weber fraction in comparison to other measures of ANS acuity. Critically, they claimed that w was problematic for many reasons: its distribution was not normal but right-skewed, its test-retest reliability was poorer than every other measure of ANS acuity, and more fundamentally, its value was still affected by the way low level visual cues were manipulated in the task. These results do not support the view that the Weber fraction is an invariable psychophysiological parameter devoid of context. In addition, the authors reported that w highly correlated with overall accuracy throughout the task. In other words, w was not more precise nor more informative than general accuracy. The advantages of using this parameter are thus disputable, yet it is commonly used and referred to in the literature as an appropriate tool to compare data sets from different published studies (e.g., in Castronovo and Göbel, 2012; Halberda et al., 2012; Geary et al., 2015; Libertus et al., 2016).

In the current meta-analysis, we aim at verifying whether Weber fractions computed from various numerical comparison tasks are stable and consistent in the literature. If this were the case, then its usage should be preferred to compare datasets from different studies. Alternatively, the observation of substantial heterogeneity in Weber fractions would be worrying for researchers and for clinicians who want to compare performance from a particular sample or from an individual to some *typical* performance.

METHODS

Article Search and Inclusion Criteria

The current meta-analysis only included peer-reviewed articles written in English and published before January 1st 2018 in any scientific journal. Following these inclusion criteria, we independently searched in the three databases *PsycINFO*, *PubMed*, and *Web of Science* for the documents that included the whole expression “*Approximate Number System*” in their title, abstract, keyword, or main body. The cross-referencing of the three searches yielded 387 unique references. We refined the search by looking within each document for any mention of the terms “*Weber*” or “*fraction*”. We gathered all matching articles and select the ones that (a) described at least one empirical study conducted on humans with no history of atypical development and that (b) explicitly reported the mean value of the Weber fraction (computed from any non-symbolic comparison task) of their sample(s). Sixty-eight publications were thereby included

in the current meta-analysis. They are further referenced in our bibliography with an asterisk. All statistical analyses were conducted on R Studio (R Core Team R., 2016).

We considered two substantial aspects affecting the evaluation of w in the current meta-analysis, in order to minimize any potential risk of bias. First, we highlighted each reported w as a function of the mean age of the participants. As noted by Halberda et al. (2012), performance—and subsequently w —is intrinsically more heterogeneous in children than in adults (see also, Siegler, 2007). It is consequently insufficient to investigate the variability of this measure within young children. For this reason, we decided to focus on young adults to get a clearer picture of the stability of w throughout the literature. Such picture is actually critical to support any claim that Weber fractions are reliable and invariable measures of ANS acuity.

Secondly, and more critically for the purpose of the current meta-analysis, the procedure used to generate stimuli—and to control for non-numerical visual cues—does not have a negligible impact on the value of the Weber fraction (Inglis and Gilmore, 2014; Clayton et al., 2015; Smets et al., 2016). In numerical comparison tasks, participants are sensitive to non-numerical dimensions, and they might base their judgments on them (see Gebuis et al., 2016), so that any systematic confound between the number and one visual property substantially affects behavior. In other words, participants might strategically use available visual information to help them to respond to the task (e.g., the larger array is likely to have more elements). Therefore, paradigms that control for various non-numerical cues at the same time lead to worse performance—and thus larger w —than methods involving the manipulation of only one dimension (Smets et al., 2015). The values of w reported in a given publication are thus not independent of the properties relative to the methodology used to acquire the data (see for instance, Dietrich et al., 2015b). Although we did not aim for the evaluation of specific influence of a given generation algorithm on participants' performance, we decided to emphasize the properties of the task—and their stimuli—that underlay every considered w . However, it should be noted that we did not consider any other methodological aspects that may affect performance (such as the duration of stimulus presentation or the range of the displayed numerical quantities, see Clayton et al., 2015; Smets et al., 2016), as they drastically fluctuated from studies to studies and were thus difficult to categorize in such meta-analysis. We describe how we categorize the dataset in the following section.

Algorithm Description and Categorization

It is worth noting that a sizeable amount of the retained publications described data collected from more than one participant sample (e.g., comparing different age groups or different methodologies, having different data points in a longitudinal setup). For this reason, we decided to consider data at the sample level, and not at the study level. We then arranged all samples by three categories.

The first category contains the typically developing human samples from publications that explicitly mention the use of the *Panamath*, which is an assessment software freely available

at www.panamath.org. *Panamath* is actually the only existing program that can be implemented with the greatest of ease to test participants or patients, and to directly obtain a performance index, as well as the computation of their Weber fraction. It is thus well known among researchers and practitioners interested in evaluating non-symbolic number abilities. Experimental paradigms of all the samples within this category thus share strong similarities due to the use of the same software. This especially includes the display of two arrays of dots with different colors (blue and yellow) at the same time (see Halberda et al., 2008; for further methodological details). Nonetheless, there may still be some dissimilarity between the experimental conditions because *Panamath* allows researchers to modify some stimuli properties at their best convenience, such as the display duration and the maximal array size. It should be noted that these adaptations are primarily intended to account for the potential youth of the subject taking the test. Anyway, we disregarded such slight modifications in the current meta-analysis, and we considered all samples assessed with *Panamath* in one category.

The second category comprises the samples from studies following Dehaene's et al. (unpublished manuscript) recommendations to construct their stimuli (from Piazza et al., 2004). The authors highlighted in their manuscript that some visual properties are inherently confounded with the number of items in an array. For instance, the picture with the largest number of items is expected to occupy the largest area and/or to possess on average the smallest elements. For this reason, they suggested using a generation algorithm designed to maintain constant one visual property across both displayed arrays, so that this dimension could not be informative to make the decision. Typically, such scripts either consider the individual item size (IIS) or the total occupied area (TOA). However, because number (N) is the multiplicative factor between these two parameters, such as $IIS \times N = TOA$, the dimension that is not kept constant across the arrays is systematically correlated with number. To overcome this limitation, Dehaene et al. (unpublished manuscript) recommended generating exactly half of the stimuli with one constant dimension, and the other half with the second unvarying parameter, so that a participant that would strategically make use of the information from one non-numerical parameter would obtain the correct answer in only 50% of the case (which is the chance level). In our meta-analysis, we labeled these programs as "One-dimensional" algorithm as they control for one visual dimension at a time. Noticeably, the *Panamath* software follows this creation rule, as it controls for one visual dimension at a time. Yet the item sizes within one given array are not constant in *Panamath*. We thus excluded this script from the second category and we only considered here studies following the half IIS/half TOA constant rule from Dehaene et al. (unpublished manuscript), without any further restriction.

In the third and last category, we considered all other studies that did not use any of the previously described generation scripts. It is noteworthy that none of these manuscripts put aside the methodological concern that many visual dimensions are inherently confounded with number. On the contrary, their generation procedures all featured their own consideration for

controlling for more than one visual parameter at a time (besides IIS and TOA, such as the length of the convex hull formed by the array or the item density). Among these procedures, one could refer to Gebuis and Reynvoet’s (2011) program that manipulates the congruity (or incongruity) of five different dimensions with number throughout the stimuli, to the paradigm of Mussolin et al. (2012) that used collections with various elements richer than single dots, and to the paradigm of DeWind et al. (2015) that disentangles the relative contribution of three orthogonal dimensions (number, spacing, and size) within participants’ performance. As these methods accounted for more than one-dimension at a time, we labeled these “Multi-dimensional” algorithms.

RESULTS

Description of the Considered Samples

Within the 68 scientific publications that were considered in the current meta-analysis, we retained 115 samples of typically developing humans. Nineteen documents that together described data from 28 samples explicitly mentioned using the *Panamath*. Thirty-six articles were included in the “One-dimensional” category, for a total of 63 typical samples that used such generation algorithm. The third category contained 15 documents that reported on 24 typical samples that used “Multi-dimensional” programs. Descriptive data of considered documents and samples are indicated in **Table 1**. The whole list is

TABLE 1 | Description of the data, as a function of the generation algorithm.

Generation algorithm	Count of documents	Count of samples	Count of participants	Mean sample age (SD)	Mean weber fraction (SD)
Panamath	19	28	1975	12.9 years (11.7)	0.29 (0.17)
One-dimensional	36	63	5230	15.1 years (10.9)	0.30 (0.18)
Multi-dimensional	15	24	882	14.8 years (9.21)	0.33 (0.16)
Total	70 ^a	115	8087	14.5 years (10.7)	0.30 (0.17)

^aOne manuscript (Smets et al., 2016) contained two studies that used two different generation algorithms, one one-dimensional and one multi-dimensional. Another document (Smets et al., 2015) reported on two one-dimensional conditions and one multi-dimensional condition. They are thus considered twice in the count of documents in Table 1. Values in italic (and in brackets) are Standard Deviations, which is specified in the column title by their common acronym as (SD).

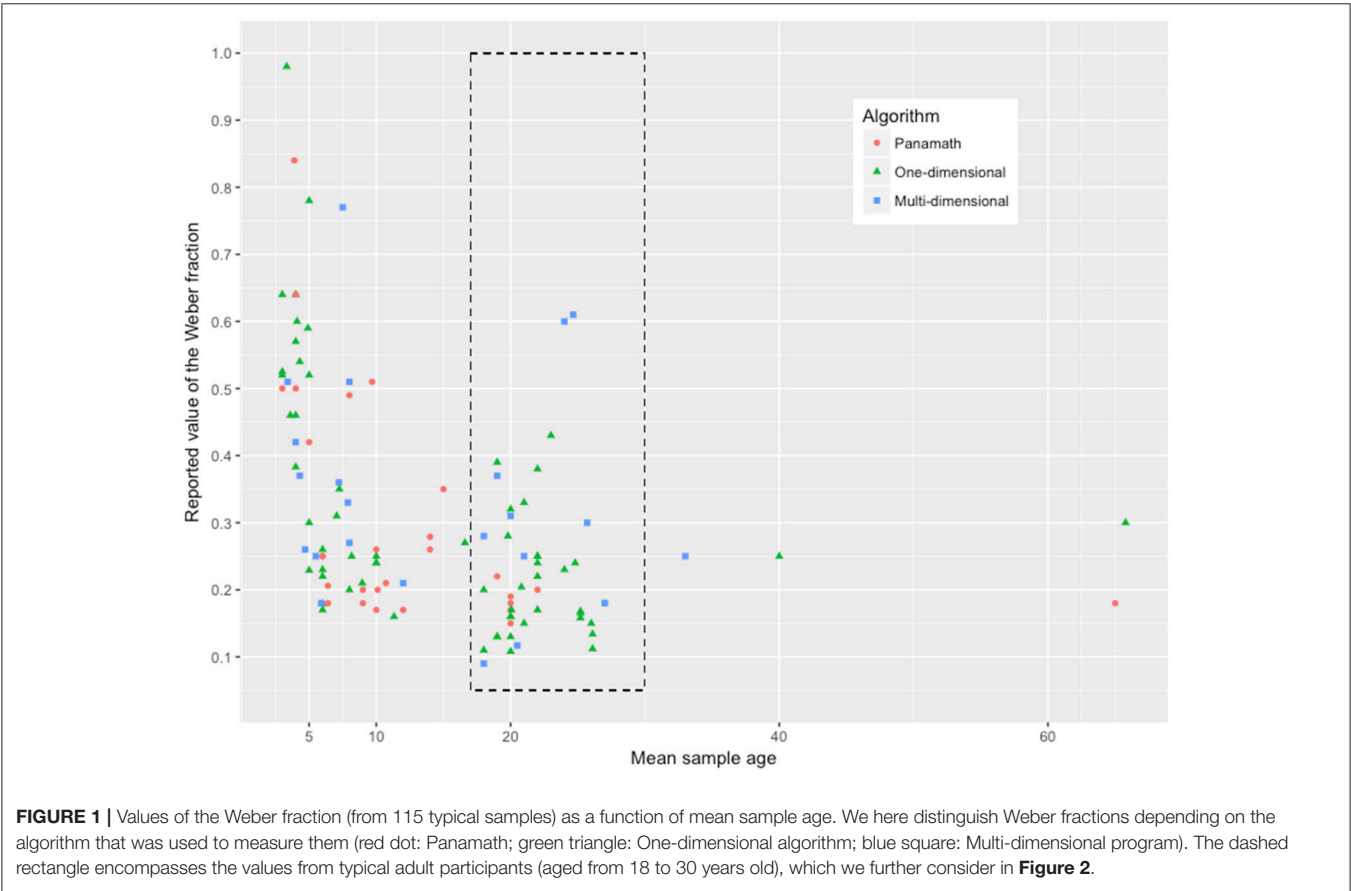


FIGURE 1 | Values of the Weber fraction (from 115 typical samples) as a function of mean sample age. We here distinguish Weber fractions depending on the algorithm that was used to measure them (red dot: Panamath; green triangle: One-dimensional algorithm; blue square: Multi-dimensional program). The dashed rectangle encompasses the values from typical adult participants (aged from 18 to 30 years old), which we further consider in **Figure 2**.

available in **Supplementary Table S1**. Overall, mean sample age was 14.5 years, 95% *CI* [12.5, 16.5], and mean Weber fraction value was 0.30, 95% *CI* [0.27, 0.34]. Two one-way analyses of variance—with Generation algorithm as group factor—revealed that both mean sample age and mean Weber fraction did not significantly differ between the three generation algorithms, both $F_{(2, 112)} < 1$. We finally conducted an ANCOVA on the value of w controlling for mean age, with Generation algorithm as the group factor; this analysis did not lead to any significant effect, $F_{(3, 111)} < 1$.

Weber Fractions in Adults

Despite the overall absence of a significant difference between the three algorithm categories in terms of Weber fractions, a closer look at **Figure 1** revealed that w means were not totally independent from mean sample age. Pearson correlation coefficient between the two variables was at $r = -0.41$, which was indeed significant, $t_{(113)} = -4.662$, $p < 0.001$. The value of the Weber fraction thus diminished when age increased, which was in line with previous findings that the noisier numerical acuity at younger age is going through some developmental changes and gradually refines over the years (until ~ 30 years, Halberda et al., 2012). Moreover, due to the inherent variability of data collected in children (Siegler, 2007), we focused our further analyses on samples of adults ranging

from 18 to 30 years, in order to be able to compare similar data.

Mean Weber fractions from these selected samples are depicted in **Figure 2A**. Data was collected from 34 documents comprising 47 typically developed adult samples. Mean sample age was 21.68, 95% *CI* [20.89, 22.46], and mean sample size was 48, 95% *CI* [35, 60]. Critically, the mean Weber fraction was 0.22, 95% *CI* [0.19, 0.26]. The latter value drastically ranged, from a minimal value of 0.09 (“congruent condition” from Smets et al., 2015) to a maximal value of 0.61 (in Dietrich et al., 2016). Even in young adult samples, which are expected to be stable, ANS acuity was thus prone to depict substantial heterogeneity.

Such variability seemed to be relative to the generation rule that was followed to create the stimuli (see **Figure 2B**). The *Panamath* lead to the smallest average value: $w = 0.18$, 95% *CI* [0.15, 0.21]. Studies with any One-dimensional algorithm observed a mean w value of 0.20, 95% *CI* [0.17, 0.24]. On the other hand, Multi-dimensional algorithms entailed the largest mean w value of 0.29, 95% *CI* [0.18, 0.41]. An unilateral Welch test of equality of means revealed that the algorithm category impacted the mean value of the Weber fraction, $F_{(2, 20.388)} = 2.768$, $p = 0.043$. Pairwise comparison tests (Bonferroni corrected) revealed that the *Panamath* and the One-dimensional category did not significantly differ from each other, $p = 0.627$; however, Multi-dimensional algorithms were significantly greater than the other two, $p = 0.024$ and $p =$

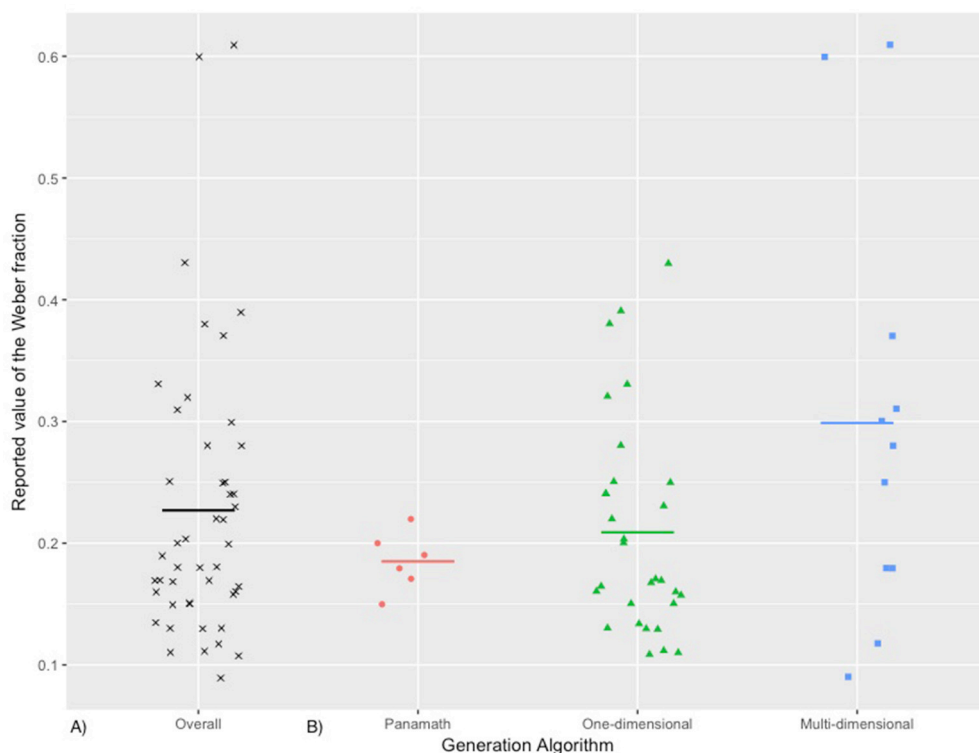


FIGURE 2 | (A) Values of the Weber fraction (from 47 samples of typical adults). **(B)** Values of the Weber fraction as a function of the algorithm that was used to measure these values (red dot: *Panamath*; green triangle: One-dimensional algorithm; blue square: Multi-dimension program). The horizontal lines depict the mean values.

0.046 respectively for the *Panamath* and the One-dimensional type. Furthermore, we conducted a Brown-Forsythe test for homogeneity of variance, and this test revealed that variance was statistically different between the algorithms, $F_{(2, 44)} = 3.965$, $p = 0.026$. This confirms that the variability of the values was different between the three categories. In other words, as depicted in **Figure 2B**, *Panamath* was less variable than the other generation scripts.

Incidentally, in the publications considered in previous analysis, there were 11 explicit reports of Pearson correlation coefficient between the Weber fraction and overall accuracy in the numerical comparison task. Reported coefficient were very high as the mean $r = 0.97$, 95% *CI* [0.96, 0.98], with a minimum r of 0.90. This is not surprising, as Weber fractions are computed from accuracy scores (Piazza et al., 2004; Halberda and Feigenson, 2008). As Inglis and Gilmore (2014) pointed out, such high correlation coefficients question whether the Weber fraction is more informative than general accuracy score, and whether the former should be preferred over the latter.

DISCUSSION

In the current meta-analysis, we highlighted that the Weber fractions computed from numerical comparison tasks are heterogeneous, even within young adult samples. This variability does not support the view that w is a stable parameter devoid of context. As many authors surmised, methodological specificities of the numerical tasks used to compute w impacted its value (e.g., Clayton et al., 2015; Smets et al., 2016), and we were able to characterize this substantial heterogeneity in the literature. As depicted in **Figures 1, 2**, the method used to generate the non-symbolic arrays substantially affected the mean and the variance of the values of w . Multi-dimensional algorithms led to larger w than the other generation programs. This is likely due to the strategic use—or the unconscious experience—of the non-numerical information that is automatically extracted in the visual cortex during the task (Gebuis et al., 2016; Leibovich et al., 2016). At this point, we want to emphasize that we did not aim for the exhaustive description of all methodological discrepancies that might affect measures of ANS acuity (see Dietrich et al., 2015b). For instance, we did not analyse the impact of the range of the quantities used in each study in their evaluation of w . In addition, the current meta-analysis did not provide any theoretical evidence that the ANS does not exist (Leibovich et al., 2016; see alternative view from Gebuis et al., 2016). Our analysis only provides evidence that w is not an invariable measure, which may explain some substantial parts in the relation (or non-relation) between ANS acuity and math ability (e.g., Price et al., 2012; Gilmore et al., 2013; Sasanguie et al., 2013).

Interestingly, studies that specifically used the *Panamath* reported homogeneous results. This suggests that the slight methodological difference that we did not consider between these studies—such as the stimulus duration or the numerical range of the arrays—did not drastically impact the measurement of w . In other words, *Panamath* studies were thus robust to small dissimilarities in evaluating w . It should be noted that our analysis

does not simply imply that the *Panamath* reliably assess ANS acuity (see Gebuis et al., 2016; for more detailed methodological considerations). Some might claim that studies from the same laboratory or that use the same exact paradigm arguably tend to show overall higher consistencies, independently from the nature of the task. That being said, our meta-analysis supports that it is possible to reliably measure the same cognitive process in similar numerical comparison tasks, which is reassuring for the literature. It is indeed essential to ascertain that different studies are assessing the same cognitive process before drawing further conclusions about ANS acuity and math ability (see Maxwell et al., 2015; for further considerations about the relevance of replication studies).

Finally, in line with Inglis and Gilmore, 2014 observation, w strongly correlated with general accuracy in the literature. It is unsurprising, as w indexes *in fine* participant accuracy throughout a numerical task. Yet accuracy is modulated by the way non-numerical visual cues are manipulated, with lower performance when multiple visual dimensions are manipulated at the same time (Smets et al., 2015, 2016). The Weber fraction thus does not provide any additional information about performance than overall accuracy does, mostly when taking a correlational perspective. As Inglis and Gilmore (2014) emphasized, one may wonder whether we should compute w at all in the future. With the exception of precise psychophysiological modeling of datasets to highlight specific contribution of numerical and non-numerical dimensions on human behavior (as in DeWind et al., 2015), we believe that most researchers and most clinicians should not bother computing w . On the contrary, emphasizing w might give the false impression of its invariability, which might incorrectly encourage direct comparison of very different datasets, whereas reporting percentages of correct responses would not favor such direct comparison. This is not trivial, as the evaluation and the training of ANS acuity both have a substantial clinical impact in the assessment and the remediation of math disability (Mazzocco et al., 2011; Park and Brannon, 2013). We are thus in disfavor of the systematic use of the Weber fraction and in favor of the consideration of normative accuracy datasets acquired from the exact same numerical comparison task.

In conclusion, the Weber fraction is an appealing measure of numerical discrimination due to its psychophysiological nature. It is a precious tool to precisely model human behavior. However, researchers and clinicians should not be unaware of its heterogeneity and its context-dependence. The algorithm used to generate the stimulus set within the task substantially affects its value and its variability. This measure is thus not directly transferable from one study to another. Researchers and practitioners should thus be extremely cautious when comparing comparison tasks.

AUTHOR CONTRIBUTIONS

MG and AV: original idea and revision; MG: data collection, data analysis, and drafting manuscript.

FUNDING

MG is funded by an internal research fund from the University of Luxembourg (IRP_2015_CAFA). AV is funded by an European Marie Skłodowska-Curie Action project (N° 799171).

REFERENCES

- *Anobile, G., Cicchini, G. M., and Burr, D. C. (2014). Separate mechanisms for perception of numerosity and density. *Psychol. Sci.* 25, 265–270. doi: 10.1177/0956797613501520
- *Attout, L., Noël, M.-P., Vossius, L., and Rousselle, L. (2017). Evidence of the impact of visuo-spatial processing on magnitude representation in 22q11.2 microdeletion syndrome. *Neuropsychologia* 99, 296–305. doi: 10.1016/j.neuropsychologia.2017.03.023
- Barth, H., La Mont, K., Lipton, J. S., Dehaene, S., Kanwisher, N., and Spelke, E. S. (2006). Non-symbolic arithmetic in adults and young children. *Cognition* 98, 199–222. doi: 10.1016/j.cognition.2004.09.011
- *Bugden, S., and Ansari, D. (2016). Probing the nature of deficits in the “Approximate Number System” in children with persistent Developmental Dyscalculia. *Dev. Sci.* 19, 817–833. doi: 10.1111/desc.12324
- Cantlon, J. F., Cordes, S., Libertus, M. E., and Brannon, E. M. (2009). Comment on “Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures”. *Science* 323:38b. doi: 10.1126/science.1156540
- *Cappelletti, M., Didino, D., Stoianov, I., and Zorzi, M. (2014). Number skills are maintained in healthy ageing. *Cogn. Psychol.* 69, 25–45. doi: 10.1016/j.cogpsych.2013.11.004
- Castronovo, J., and Göbel, S. M. (2012). Impact of high mathematics education on the number sense. *PLoS ONE* 7:e33832. doi: 10.1371/journal.pone.0033832
- *Chesney, D., Bjälkebring, P., and Peters, E. (2015). How to estimate how well people estimate: evaluating measures of individual differences in the approximate number system. *Attention Percept. Psychophys.* 77, 2781–2802. doi: 10.3758/s13414-015-0974-6
- *Chu, F. W., vanMarle, K., and Geary, D. C. (2013). Quantitative deficits of preschool children at risk for mathematical learning disability. *Front. Psychol.* 4:195. doi: 10.3389/fpsyg.2013.00195
- Cicchini, G. M., Anobile, G., and Burr, D. C. (2014). Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proc. Natl. Acad. Sci.* 111, 7867–7872. doi: 10.1073/pnas.1402785111
- Clayton, S., Gilmore, C., and Inglis, M. (2015). Dot comparison stimuli are not all alike: the effect of different visual controls on ANS measurement. *Acta Psychol.* 161, 177–184. doi: 10.1016/j.actpsy.2015.09.007
- Clayton, S., Inglis, M., and Gilmore, C. (2018). Developmental differences in approaches to nonsymbolic comparison tasks. *Quart. J. Exp. Psychol.* 1:1747021818755296. doi: 10.1177/1747021818755296
- Dehaene, S. (1997). *The Number Sense: How the Mind creates Mathematics*. New York, NY: Oxford University Press.
- Dehaene, S. (2003). The neural basis of the Weber–Fechner law: a logarithmic mental number line. *Trends Cogn. Sci.* 7, 145–147. doi: 10.1016/S1364-6613(03)00055-X
- DeWind, N. K., Adams, G. K., Platt, M. L., and Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition* 142, 247–265. doi: 10.1016/j.cognition.2015.05.016
- *DeWind, N. K., and Brannon, E. M. (2016). Significant inter-test reliability across approximate number system assessments. *Front. Psychol.* 7:310. doi: 10.3389/fpsyg.2016.00310
- *Dietrich, J. F., Huber, S., Klein, E., Willmes, K., Pixner, S., and Moeller, K. (2016). A systematic investigation of accuracy and response time based measures used to index ANS acuity. *PLoS ONE* 11:e0163076. doi: 10.1371/journal.pone.0163076
- *Dietrich, J. F., Huber, S., Moeller, K., and Klein, E. (2015a). The influence of math anxiety on symbolic and non-symbolic magnitude processing. *Front. Psychol.* 6:1621. doi: 10.3389/fpsyg.2015.01621

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01694/full#supplementary-material>

- Dietrich, J. F., Huber, S., and Nuerk, H.-C. (2015b). Methodological aspects to be considered when measuring the approximate number system (ANS): a research review. *Front. Psychol.* 6:295. doi: 10.3389/fpsyg.2015.00295
- *Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53–72. doi: 10.1016/j.jecp.2014.01.013
- Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002
- Gallistel, C. R., and Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends Cogn. Sci.* 4, 59–65. doi: 10.1016/S1364-6613(99)01424-2
- *Geary, D. C., Hoard, M. K., Nugent, L., and Rouders, J. N. (2015). Individual differences in algebraic cognition: relation to the approximate number and semantic memory systems. *J. Exp. Child Psychol.* 140, 211–227. doi: 10.1016/j.jecp.2015.07.010
- Gebuis, T., Cohen Kadosh, R., and Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: a critical review. *Acta Psychol.* 171, 17–35. doi: 10.1016/j.actpsy.2016.09.003
- Gebuis, T., and Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behav. Res. Methods* 43, 981–986. doi: 10.3758/s13428-011-0097-5
- Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., et al. Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS ONE* 8:e67374. doi: 10.1371/journal.pone.0067374
- *Gilmore, C., Attridge, N., and Inglis, M. (2011). Measuring the approximate number system. *Q. J. Exp. Psychol.* 64, 2099–2109. doi: 10.1080/17470218.2011.574710
- *Gomez, A., Piazza, M., Jobert, A., Dehaene-Lambertz, G., Dehaene, S., and Huron, C. (2015). Mathematical difficulties in developmental coordination disorder: symbolic and nonsymbolic number processing. *Res. Dev. Disabilities* 43–44, 167–178. doi: 10.1016/j.ridd.2015.06.011
- *Guillaume, M., Gevers, W., and Content, A. (2016). Assessing the approximate number system: no relation between numerical comparison and estimation tasks. *Psychol. Res.* 80, 248–258. doi: 10.1007/s00426-015-0657-x
- *Guillaume, M., Nys, J., Mussolin, C., and Content, A. (2013). Differences in the acuity of the Approximate Number System in adults: the effect of mathematical ability. *Acta Psychol.* 144, 506–512. doi: 10.1016/j.actpsy.2013.09.001
- *Halberda, J., and Feigenson, L. (2008). Developmental change in the acuity of the “number sense”: the approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Dev. Psychol.* 44, 1457–1465. doi: 10.1037/a0012682
- Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11116–11120. doi: 10.1073/pnas.1200196109
- *Halberda, J., Mazzocco, M. M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, 455, 665–668. doi: 10.1038/nature07246
- *Hellgren, K., Halberda, J., Forsman, L., Ådén, U., and Libertus, M. (2013). Compromised approximate number system acuity in school-aged children born extremely preterm. *Dev. Med. Child Neurol.* 55, 1070–1071. doi: 10.1111/dmcn.12206
- *Inglis, M., and Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychol.* 145, 147–155. doi: 10.1016/j.actpsy.2013.11.009
- *Jang, S., and Cho, S. (2016). The acuity for numerosity (but not continuous magnitude) discrimination correlates with quantitative problem solving but not routinized arithmetic. *Curr. Psychol.* 35, 44–56. doi: 10.1007/s12144-015-9354-6

- *Jones, S. M., Pearson, J., DeWind, N. K., Paulsen, D., Tenekedjieva, A.-M., and Brannon, E. M. (2014). Lemurs and macaques show similar numerical sensitivity. *Anim. Cogn.* 17, 503–515. doi: 10.1007/s10071-013-0682-3
- Julio-Costa, A., Starling-Alves, I., Lopes-Silva, J. B., Wood, G., and Haase, V. G. (2015). Stable measures of number sense accuracy in math learning disability: Is it time to proceed from basic science to clinical application? *Psych J.* 4, 218–225. doi: 10.1002/pchj.114
- *Khanum, S., Hanif, R., Spelke, E. S., Berteletti, I., and Hyde, D. C. (2016). Effects of non-symbolic approximate number practice on symbolic numerical abilities in pakistani children. *PLoS ONE* 11:e0164436. doi: 10.1371/journal.pone.0164436
- *Knops, A., Dehaene, S., Berteletti, I., and Zorzi, M. (2014). Can approximate mental calculation account for operational momentum in addition and subtraction? *Quart. J. Exp. Psychol.* 67, 1541–1556. doi: 10.1080/17470218.2014.890234
- Leibovich, T., Katzin, N., Harel, M., and Henik, A. (2016). From “sense of number” to “sense of magnitude”: The role of continuous magnitudes in numerical cognition. *Behav. Brain Sci.* 40, 1–62. doi: 10.1017/S0140525X16000960
- *Libertus, M. E., Feigenson, L., and Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability: approximate number system and math abilities. *Dev. Sci.* 14, 1292–1300. doi: 10.1111/j.1467-7687.2011.01080.x
- *Libertus, M. E., Feigenson, L., and Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learn. Individ. Differ.* 25, 126–133. doi: 10.1016/j.lindif.2013.02.001
- *Libertus, M. E., Feigenson, L., Halberda, J., and Landau, B. (2014). Understanding the mapping between numerical approximation and number words: evidence from Williams syndrome and typical development. *Dev. Sci.* 17, 905–919. doi: 10.1111/desc.12154
- *Libertus, M. E., Odic, D., Feigenson, L., and Halberda, J. (2016). The precision of mapping between number words and the approximate number system predicts children’s formal math abilities. *J. Exp. Child Psychol.* 150, 207–226. doi: 10.1016/j.jecp.2016.06.003
- *Libertus, M. E., Odic, D., and Halberda, J. (2012). Intuitive sense of number correlates with math scores on college-entrance examination. *Acta Psychol.* 141, 373–379. doi: 10.1016/j.actpsy.2012.09.009
- *Lindskog, M., Winman, A., and Juslin, P. (2014). The association between higher education and approximate number system acuity. *Front. Psychol.* 5:462. doi: 10.3389/fpsyg.2014.00462
- *Lindskog, M., Winman, A., Juslin, P., and Poom, L. (2013). Measuring acuity of the approximate number system reliably and validly: the evaluation of an adaptive test procedure. *Front. Psychol.* 4:510. doi: 10.3389/fpsyg.2013.00510
- *Lindskog, M., Winman, A., and Poom, L. (2016). Arithmetic Training does not improve approximate number system acuity. *Front. Psychol.* 7:1634. doi: 10.3389/fpsyg.2016.01634
- *Liu, Y. (2017). Fraction magnitude understanding and its unique role in predicting general mathematics achievement at two early stages of fraction instruction. *Br. J. Educ. Psychol.* 88:345–362. doi: 10.1111/bjep.12182
- *Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., and Lindberg, S. (2013). Developmental changes in the association between approximate number representations and addition skills in elementary school children. *Front. Psychol.* 4:783. doi: 10.3389/fpsyg.2013.00783
- *Lourenco, S. F., Bonny, J. W., Fernandez, E. P., and Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proc. Natl. Acad. Sci.* 109, 18737–18742. doi: 10.1073/pnas.1207212109
- *Lyons, I. M., and Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition* 121, 256–261. doi: 10.1016/j.cognition.2011.07.009
- Maxwell, S. E., Lau, M. Y., and Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does “failure to replicate” really mean? *Am. Psychol.* 70, 487–498. doi: 10.1037/a0039400
- *Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (Dyscalculia): impaired numerical acuity contributes to MLD. *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x
- *Mussolin, C., Nys, J., Content, A., and Leybaert, J. (2014). Symbolic number abilities predict later approximate number system acuity in preschool children. *PLoS ONE* 9:e91839. doi: 10.1371/journal.pone.0091839
- *Mussolin, C., Nys, J., Leybaert, J., and Content, A. (2012). Relationships between approximate number system acuity and early symbolic number abilities. *Trends Neurosci. Educ.* 1, 21–31. doi: 10.1016/j.tine.2012.09.003
- *Norris, J. E., and Castronovo, J. (2016). Dot Display Affects approximate number system acuity and relationships with mathematical achievement and inhibitory control. *PLoS ONE* 11:e0155543. doi: 10.1371/journal.pone.0155543
- *Norris, J. E., McGeown, W. J., Guerrini, C., and Castronovo, J. (2015). Aging and the number sense: preserved basic non-symbolic numerical processing and enhanced basic symbolic processing. *Front. Psychol.* 6:999. doi: 10.3389/fpsyg.2015.00999
- *Núñez, R. E. (2017). Is there really an evolved capacity for number? *Trends Cogn. Sci.* 21, 409–424. doi: 10.1016/j.tics.2017.03.005
- *Odic, D. (2017). Children’s intuitive sense of number develops independently of their perception of area, density, length, and time. *Dev. Sci.* 21:e12533. doi: 10.1111/desc.12533
- *Odic, D., and Halberda, J. (2015). Eye movements reveal distinct encoding patterns for number and cumulative surface area in random dot arrays. *J. Vis.* 15:5. doi: 10.1167/15.15.5
- *Odic, D., Hock, H., and Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *J. Exp. Psychol.* 143, 255–265. doi: 10.1037/a0030825
- *Odic, D., Libertus, M. E., Feigenson, L., and Halberda, J. (2013a). Developmental change in the acuity of approximate number and area representations. *Dev. Psychol.* 49, 1103–1112. doi: 10.1037/a0029472
- *Odic, D., Pietroski, P., Hunter, T., Lidz, J., and Halberda, J. (2013b). Young children’s understanding of “more?” and discrimination of number and surface area. *J. Exp. Psychol.* 39, 451–461. doi: 10.1037/a0028874
- *Oliveira, L. F. S., Santos, A. O., Vianna, G. S., Di Nanno, C. Q. M. S., Giacheti, C. M., Carvalho, M. R. S., et al. Haase, V. G. (2014). Impaired acuity of the approximate number system in 22q11.2 microdeletion syndrome. *Psychol. Neurosci.* 7, 151–158. doi: 10.3922/j.psns.2014.02.04
- *Olsson, L., Östergren, R., and Träff, U. (2016). Developmental dyscalculia: a deficit in the approximate number system or an access deficit? *Cogn. Dev.* 39, 154–167. doi: 10.1016/j.cogdev.2016.04.006
- Park, J., and Brannon, E. M. (2013). Training the Approximate number system improves math proficiency. *Psychol. Sci.* 24, 2013–2019. doi: 10.1177/0956797613482944
- *Park, J., and Starns, J. J. (2015). The approximate number system acuity redefined: a diffusion model approach. *Front. Psychol.* 6:1955. doi: 10.3389/fpsyg.2015.01955
- *Patalano, A. L., Saltiel, J. R., Machlin, L., and Barth, H. (2015). The role of numeracy and approximate number system acuity in predicting value and probability distortion. *Psychon. Bull. Rev.* 22, 1820–1829. doi: 10.3758/s13423-015-0849-9
- *Piantadosi, S. T. (2016). A rational analysis of the approximate number system. *Psychon. Bull. Rev.* 23, 877–886. doi: 10.3758/s13423-015-0963-8
- *Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012
- Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* 44, 547–555. doi: 10.1016/j.neuron.2004.10.014
- Piazza, M., Pica, P., Izard, V., Spelke, E. S., and Dehaene, S. (2013). Education enhances the acuity of the nonverbal approximate number system. *Psychol. Sci.* 24, 1037–1043. doi: 10.1177/0956797612464057
- Pica, P., Lemer, C., Izard, V., and Dehaene, S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science*, 306, 499–503. doi: 10.1126/science.1102085
- *Pietroski, P., Lidz, J., Hunter, T., and Halberda, J. (2009). The meaning of “most”: Semantics, numerosity and psychology. *Mind Lang.* 24, 554–585. doi: 10.1111/j.1468-0017.2009.01374.x
- *Pinheiro-Chagas, P., Wood, G., Knops, A., Krinzinger, H., Lonnemann, J., Starling-Alves, I., et al. (2014). In how many ways is the approximate number system associated with exact calculation? *PLoS ONE* 9:e111155. doi: 10.1371/journal.pone.0111155

- Platt, J. R., and Johnson, D. M. (1971). Localization of position within a homogeneous behavior chain: effects of error contingencies. *Learn. Motiv.* 2, 386–414.
- *Price, G. R., Palmer, D., Battista, C., and Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychol.* 140, 50–57. doi: 10.1016/j.actpsy.2012.02.008
- R Core Team R. (2016). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. Available online at: <https://www.R-project.org>
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: what underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418–431. doi: 10.1016/j.jecp.2012.10.012
- *Shusterman, A., Slusser, E., Halberda, J., and Odic, D. (2016). Acquisition of the cardinal principle coincides with improvement in approximate number system acuity in preschoolers. *PLoS ONE* 11:e0153072. doi: 10.1371/journal.pone.0153072
- Siegler, R. S. (2007). Cognitive variability. *Dev. Sci.* 10, 104–109. doi: 10.1111/j.1467-7687.2007.00571.x
- *Skagerlund, K., and Träff, U. (2014). Development of magnitude processing in children with developmental dyscalculia: space, time, and number. *Front. Psychol.* 5:675. doi: 10.3389/fpsyg.2014.00675
- *Skagerlund, K., and Träff, U. (2016). Processing of space, time, and number contributes to mathematical abilities above and beyond domain-general cognitive abilities. *J. Exp. Child Psychol.* 143, 85–101. doi: 10.1016/j.jecp.2015.10.016
- *Smets, K., Gebuis, T., Defever, E., and Reynvoet, B. (2014). Concurrent validity of approximate number sense tasks in adults and children. *Acta Psychol.* 150, 120–128. doi: 10.1016/j.actpsy.2014.05.001
- *Smets, K., Moors, P., and Reynvoet, B. (2016). Effects of presentation type and visual control in numerosity discrimination: implications for number processing? *Front. Psychol.* 7:66. doi: 10.3389/fpsyg.2016.00066
- *Smets, K., Sasanguie, D., Szűcs, D., and Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *J. Cogn. Psychol.* 27, 310–325. doi: 10.1080/20445911.2014.996568
- *Starr, A., Libertus, M. E., and Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proc. Natl. Acad. Sci.* 110, 18116–18120. doi: 10.1073/pnas.1302751110
- Stevens, S. S. (1957). On the psychophysical law. *Psychol. Rev.* 64:153. doi: 10.1037/h0046162
- *Szűcs, D., Nobes, A., Devine, A., Gabriel, F. C., and Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Front. Psychol.* 4:444. doi: 10.3389/fpsyg.2013.00444
- *Tinelli, F., Anobile, G., Gori, M., Aagten-Murphy, D., Bartoli, M., Burr, D. C., et al. Concetta Morrone, M. (2015). Time, number and attention in very low birth weight children. *Neuropsychologia* 73, 60–69. doi: 10.1016/j.neuropsychologia.2015.04.016
- *Tosto, M. G., Petrill, S. A., Halberda, J., Trzaskowski, M., Tikhomirova, T. N., Bogdanova, O. Y., et al. (2014). Why do we differ in number sense? Evidence from a genetically sensitive investigation. *Intelligence* 43, 35–46. doi: 10.1016/j.intell.2013.12.007
- *van den Berg, R., Lindskog, M., Poom, L., and Winman, A. (2017). Recent is more: a negative time-order effect in nonsymbolic numerical judgment. *J. Exp. Psychol.* 43, 1084–1097. doi: 10.1037/xhp0000387
- Van Oeffelen, M. P., and Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Percept. Psychophys.* 32, 163–170. doi: 10.3758/BF03204275
- *vanMarle, K., Chu, F. W., Li, Y., and Geary, D. C. (2014). Acuity of the approximate number system and preschoolers? Quantitative development. *Dev. Sci.* 17, 492–505. doi: 10.1111/desc.12143
- *Whalen, J., Gallistel, C. R., and Gelman, R. (1999). Nonverbal counting in humans: the psychophysics of number representation. *Psychol. Sci.* 10, 130–137.
- *Winman, A., Juslin, P., Lindskog, M., Nilsson, H., and Kerimi, N. (2014). The role of ANS acuity and numeracy for the calibration and the coherence of subjective probability judgments. *Front. Psychol.* 5:851. doi: 10.3389/fpsyg.2014.00851
- *Wong, T. T., Ho, C. S.-H., and Tang, J. (2014). Identification of children with mathematics learning disabilities (MLDs) using latent class growth analysis. *Res. Dev. Disabil.* 35, 2906–2920. doi: 10.1016/j.ridd.2014.07.015
- *Wong, T. T.-Y., Ho, C. S.-H., and Tang, J. (2016). The relation between ANS and symbolic arithmetic skills: the mediating role of number-numerosity mappings. *Contemp. Educ. Psychol.* 46, 208–217. doi: 10.1016/j.cedpsych.2016.06.003

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Guillaume and Van Rinsveld. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

*Publications included in the meta-analysis are referenced with an asterisk.



The Acuity and Manipulability of the ANS Have Separable Influences on Preschoolers' Symbolic Math Achievement

Ariel Starr^{1*}, Rachel C. Tomlinson² and Elizabeth M. Brannon³

¹ Department of Psychology, University of California, Berkeley, Berkeley, CA, United States, ² Department of Psychology, University of Michigan, Ann Arbor, MI, United States, ³ Department of Psychology, University of Pennsylvania, Philadelphia, PA, United States

OPEN ACCESS

Edited by:

Xinlin Zhou,
Beijing Normal University, China

Reviewed by:

Melissa M. Kibbe,
Boston University, United States
Maciej Haman,
University of Warsaw, Poland

*Correspondence:

Ariel Starr
arielstarr@berkeley.edu

Specialty section:

This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

Received: 15 May 2018

Accepted: 28 November 2018

Published: 11 December 2018

Citation:

Starr A, Tomlinson RC and
Brannon EM (2018) The Acuity
and Manipulability of the ANS Have
Separable Influences on
Preschoolers' Symbolic Math
Achievement. *Front. Psychol.* 9:2554.
doi: 10.3389/fpsyg.2018.02554

The approximate number system (ANS) is widely considered to be a foundation for the acquisition of uniquely human symbolic numerical capabilities. However, the mechanism by which the ANS may support symbolic number representations and mathematical thought remains poorly understood. In the present study, we investigated two pathways by which the ANS may influence early math abilities: variability in the acuity of the ANS representations, and children's ability to manipulate ANS representations. We assessed the relation between 4-year-old children's performance on a non-symbolic numerical comparison task, a non-symbolic approximate addition task, and a standardized symbolic math assessment. Our results indicate that ANS acuity and ANS manipulability each contribute unique variance to preschooler's early math achievement, and this result holds after controlling for both IQ and executive functions. These findings suggest that there are multiple routes by which the ANS influences math achievement. Therefore, interventions that target both the precision and manipulability of the ANS may prove to be more beneficial for improving symbolic math skills compared to interventions that target only one of these factors.

Keywords: approximate number system, numerical cognition, math cognition, cognitive development, symbolic math

INTRODUCTION

Math ability when a child first enters schooling is the strongest predictor of later math and overall academic achievement (Duncan et al., 2007). However, there is variation in math ability across the population, and such variation is already present even before children first begin formal schooling (e.g., Libertus et al., 2011; Mazzocco et al., 2011; vanMarle et al., 2014). Many cognitive and socioeconomic factors are known to contribute to individual differences in math achievement. One of these factors is an evolutionarily ancient system for representing approximate quantities. Although humans use linguistic symbols to represent number, we also possess a system for representing number in an approximate, non-symbolic fashion. This system, termed the approximate number system (ANS), emerges independent of exposure to language or formal schooling and is present in a wide variety of non-human species, and emerges early in human development (e.g., Gallistel and Gelman, 1992; Dehaene, 1997; Hubbard et al., 2008).

The ANS is frequently hypothesized to be a cognitive foundation for symbolic math abilities. Lending support to this view is the finding that the acuity of the ANS, typically measured by an individual's ability to compare two arrays of dots, correlates with symbolic math achievement throughout the lifespan (see Chen and Li, 2013 for review; Fazio et al., 2014; Schneider et al., 2017). Importantly, ANS acuity prior to the beginning of formal math instruction is predictive of later math achievement (Mazzocco et al., 2011; Libertus et al., 2013; Starr et al., 2013b; vanMarle et al., 2014). These studies suggest that the precision of approximate number representations may contribute to children's acquisition of symbolic math principles and influence symbolic math performance throughout the lifespan.

Although many studies have focused on the link between ANS acuity and math achievement, relatively less attention has been paid to children's ability to manipulate approximate numerical quantities. Beyond simply representing quantities, the ANS enables infants (McCrink and Wynn, 2004), preschoolers (Barth et al., 2005, 2006; Gilmore et al., 2010), and monkeys (Cantlon and Brannon, 2007) to perform approximate arithmetic operations without the use of symbols or formal training. The ANS has even been shown to contribute to algebraic problem solving in preschool-aged children (Kibbe and Feigenson, 2015). Therefore, the manipulability of the ANS may form a basis for the acquisition the basic arithmetic principles that underlie symbolic math. In support of this view, children's approximate arithmetic performance at the beginning of kindergarten is predictive of their symbolic math achievement at the end of the academic year (Gilmore et al., 2010). Furthermore, practicing non-symbolic arithmetic in both preschool-aged children and adults leads to improvements in their symbolic arithmetic performance (Park and Brannon, 2013, 2014; Hyde et al., 2014; Park et al., 2016). Therefore, children who are more adept at manipulating approximate quantities in arithmetic operations may also be more adept at symbolic arithmetic because of the overlap in cognitive processes required by both forms of arithmetic. As a result of this overlap, it may be not only the *precision* of ANS representations that influences symbolic math achievement but also the *manipulability* of ANS representations.

However, though previous work suggests that the precision and manipulability of the ANS both contribute to symbolic math achievement, it is currently unknown whether these are separable factors. In other words, do children with more precise ANS representations necessarily also more adept at manipulating approximate quantities in arithmetic operations? If this is the case, then we would expect ANS manipulability to mediate the relation between ANS acuity and symbolic math achievement. Alternatively, if ANS acuity and manipulability are distinct, we would expect both factors to contribute unique variance to children's early symbolic math performance.

In the present research, we explicitly tested how ANS acuity and manipulability each contribute to symbolic math achievement in preschool-aged children. We focused on preschool-aged children because they have not yet started formal schooling, so they have not yet been exposed to formal symbolic math education. Thus, we could assess how different aspects of children's intuitive sense of number relate to their symbolic

math proficiency. Children were tested with a non-symbolic numerical comparison task to assess ANS acuity, a non-symbolic approximate addition task to assess ANS manipulability, and a standardized symbolic math test. In addition, children performed a general IQ test and a subset of children performed an executive functions task¹ in order to control for domain-general factors that also contribute to math achievement.

MATERIALS AND METHODS

Participants

One hundred and seventy children participated in this experiment (mean age: 4.59 years, range: 4.48–4.90 years; 89 female). Of these, 145 children completed the non-symbolic numerical comparison, non-symbolic addition, symbolic math, and IQ assessments, and 75 of those children additionally completed the executive functions task. Twenty-five children did not complete one or more of the primary tasks of interest and were therefore excluded from all analyses. Participants were recruited as part of a larger longitudinal studying tracking the development of numerical cognition from infancy into the preschool years. Data was collected between October 2011 and July 2015, and data collection was stopped when the lab moved to a new institution out of state.

Procedure

Children were tested in two separate sessions each lasting less than 1 h. During the first visit, children completed the symbolic math assessment, one session of the non-symbolic number comparison task, and the executive functions task. During the second visit, children completed the IQ assessment, a second session of the non-symbolic number comparison task, and a non-symbolic approximate arithmetic task. All children were tested individually in a quiet room, and the order of the tasks within each session was counterbalanced across participants. At each visit, parents gave written consent to a protocol approved by the local Institutional Review Board. Parents were compensated monetarily and children received a small toy.

Non-symbolic Numerical Comparison Task

On each trial, a touchscreen computer displayed two squares (8 cm × 9.5 cm) containing arrays of dots. Children were instructed to touch the square that contained more dots and to make this choice without counting. Arrays contained between 4 and 14 dots, and the numerical ratio between the arrays was 1:2, 2:3, 3:4, or 6:7. To control for non-numerical perceptual cues, the parameters of the arrays varied such that the smaller and larger numerical array each had the larger cumulative surface area on 50% of trials. All of the dots within a single array were homogenous in element size and color, and the color of each array varied randomly from trial to trial. Differential audio-visual feedback was provided after each trial, and children

¹The executive functions task was added to the battery after a 2013 paper (Gilmore et al., 2013) suggested that the link between ANS acuity and symbolic math achievement may be a by product of the link between inhibitory control and symbolic math achievement.

received a small sticker for each correct response to keep them engaged. Children performed practice trials until they made three consecutive correct responses or completed a maximum of ten trials. Children were tested with 60 trials in each session for a total of 120 trials at each time point. Each child's ANS acuity was estimated using a psychophysical modeling technique (e.g., Halberda and Feigenson, 2008; Piazza et al., 2010) to calculate a Weber fraction (w) based on performance in the non-symbolic numerical comparison task. The resulting value of w represents the noise in each participant's internal ANS representations, such that lower values of w correspond to less noise (i.e., higher ANS acuity).

Non-symbolic Approximate Addition Task

This task was adapted from Cantlon and Brannon (2007). On each trial, children viewed an animation that consisted of an array of dots moving behind an occluding box, followed by a second array moving behind the same occluder (**Figure 1**). This animated arithmetic sequence lasted a total of 2000 ms. Children then saw two squares containing arrays of dots and were instructed to touch the array that contained the same number of dots as had moved behind the occluder box. The choice arrays remained on the screen until a decision was made. Correct and incorrect values differed by a 1:2 or 1:4 ratio. The specific problems presented were: $1+1 = 2$, 4, or 8; $2+2 = 2$, 4, or 8; $4+4 = 2$, 4, or 8. Individual dot size varied across arrays but was homogenous within each array. Differential audiovisual feedback was provided after each trial, and children were rewarded with a small sticker for correct responses. Children performed practice trials until they made three consecutive correct responses or completed a maximum of ten trials. Children then completed a total of 42 test trials.

Executive Functions Task

The Day/Night task (Gerstadt et al., 1994) was used to assess executive functions. This task requires children to remember the relevant rule and to inhibit a prepotent verbal response. In the warm-up version, children were shown a card containing 16 sun and moon pictures in a pseudo-random order and instructed to say "day" for the sun pictures and "night" for the moon pictures as quickly as possible. Next, children were told they were going to play a silly version of the game that required saying the opposite picture names ("day" for the moon picture and "night" for the

sun picture). They were then shown a new card with 16 sun and moon pictures and instructed to say the opposite picture names as quickly as possible without making mistakes. The total time and number of errors were combined into a single efficiency score (number of correct responses divided by total time).

Standardized Assessments

Children's mathematical ability was assessed with the Test of Early Mathematics Ability (TEMA-3) (Ginsburg and Baroody, 2003), which consists of a series of verbally administered questions that assess age-appropriate counting ability, number-comparison facility, numeral literacy, and basic calculation skills. To assess general intelligence, children completed the two verbal (Guess What and Verbal Reasoning) and the two non-verbal subtests (Odd-Item Out and What's Missing) of the Reynolds Intellectual Assessment Scales (RIASs) (Reynolds and Kamphaus, 2003). The verbal subtests are oral assessments of verbal knowledge and reasoning. The non-verbal subtests are visuospatial assessments of reasoning, spatial ability and general knowledge. The scores on these four subtests were combined to create a composite IQ score for each child.

RESULTS

Descriptive statistics and a correlation table for all measures of interest can be found in **Tables 1, 2**. The complete dataset can be found in the Supplementary Material.

Preliminary Analyses

First we performed planned paired t -tests to confirm that participants' performance on the approximate addition task was modulated by ratio. Planned paired t -tests confirmed that children were both more accurate and responded more quickly on the 1:4 ratio trials compared to the 1:2 ratio trials in the approximate addition task [accuracy: $t(144) = 8.63$, $p < 0.001$; RT: $t(144) = -4.90$, $p < 0.001$], which suggests that this task engaged the ANS.

Regression Analyses

In the first series of analyses, we used multiple regression models to investigate the unique variance contributed by each of our measures of interest (**Table 3**). The first model (Model 1) examined the variance in symbolic math achievement predicted by ANS acuity (indexed by w), ANS manipulability (indexed by approximate addition performance), and IQ. This model revealed

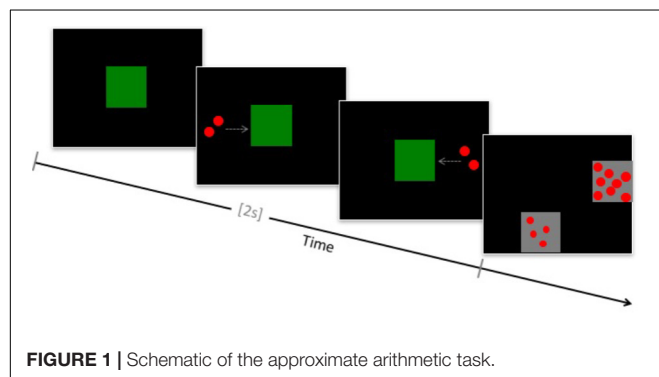


TABLE 1 | Descriptive statistics for all variables of interest.

Task	Measure	Mean (SD)
Non-symbolic numerical comparison	Accuracy (% correct)	79.45 (8.32)
	Weber fraction (w)	0.31 (0.15)
TEMA-3 (math achievement test)	Standardized score	111.57 (12.99)
RIAS (IQ test)	Standardized score	128.94 (16.15)
Day/Night (executive functions task)	Efficiency score	0.51 (0.26)
Approximate arithmetic	Accuracy (% correct)	77.27 (13.63)

TABLE 2 | Correlation matrix of Pearson *r* values for all variables of interest.

	w	Symbolic math	IQ	Approximate addition	Executive functions
w	—	−0.27	−0.16	−0.22	−0.16
Symbolic math	−0.27	—	0.42	0.32	0.33
IQ	−0.16	0.42	—	0.21	0.18
Approximate addition	−0.22	0.32	0.21	—	0.40
Executive functions	−0.16	0.33	0.18	0.40	—

TABLE 3 | Regression models predicting symbolic math achievement.

	Model 1	Model 2
R^2	0.247	0.383
<i>F</i> -statistics	$F(3,141) = 16.72$	$F(4,70) = 12.48$
<i>p</i> -statistics	$p < 0.001$	$p < 0.001$
<i>N</i>	145	75

Predictor	β_{Adjusted}	<i>p</i>	β_{Adjusted}	<i>p</i>
ANS acuity	−0.179	0.018	−0.239	0.014
ANS manipulability	0.237	0.002	0.267	0.011
IQ	0.325	0.332	<0.001	
Executive functions	—	—	0.196	0.054

that all factors contributed significant variance ($\beta_w = -0.24$, $p < 0.05$, $\beta_{\text{ApproxAdd}} = 0.27$, $p < 0.005$, $\beta_{\text{IQ}} = 0.33$, $p < 0.001$; all betas are standardized). We next ran a second model that included the executive functions task for the subset of participants who completed this task (Model 2). In this model, the original predictors all remained significant, but the executive functions task did not explain significant additional variance ($\beta_w = -0.24$, $p < 0.05$, $\beta_{\text{ApproxAdd}} = 0.27$, $p < 0.05$, $\beta_{\text{IQ}} = 0.33$, $p < 0.001$, $\beta_{\text{EF}} = 0.20$, $p = 0.054$). These analyses suggest that the acuity and manipulability of the ANS each contribute unique variance to preschooler's early symbolic math skills that is not accounted for by IQ or executive functions (Figure 2).

Mediation Analyses

Next we used structural equation modeling to determine whether the relation between ANS acuity and symbolic math achievement is mediated by ANS manipulability (Figure 3). This method enables us to directly test which portion of the relation between ANS acuity and symbolic math can be accounted for by ANS manipulability. The mediation analysis was performed using the lavaan package in R (Rosseel, 2012). The results of the mediation analyses indicate that the direct effect ($c' = -19.13$, $SE = 6.2$, $p < 0.005$) is significant whereas the indirect effect is not ($ab = -5.02$, $SE = 2.63$, $p = 0.056$). Because the direct effect remains significant after accounting for the variance contributed by the mediator and the mediation path is not significant, this suggests that ANS manipulability does not mediate the relation between ANS acuity and symbolic math achievement. Rather, ANS acuity and ANS manipulability are each making independent contributions to symbolic math achievement in preschoolers.

We also tested whether executive functions mediate the relation between approximate arithmetic performance and

symbolic math. This model indicated that both the direct effect ($c' = 13.71$, $SE = 5.43$, $p = 0.01$) and the indirect effect are significant ($ab = 6.34$, $SE = 2.72$, $p = 0.02$). Because the direct effect from approximate addition to symbolic math achievement remains significant after accounting for executive functions, this result suggests that executive functions do not fully mediate the relation between approximate arithmetic performance and math. Together, the results of these mediation analyses are consistent with the multiple regression analyses in suggesting that approximate arithmetic is contributing unique variance to children's symbolic math achievement that is not shared with ANS acuity or executive functions.

DISCUSSION

The goal of the present research was to investigate the mechanisms by which approximate number representations contribute to preschoolers' emerging symbolic math capabilities. Consistent with previous studies, we found that individual differences in the precision of the ANS are related to symbolic math achievement in preschool-aged children (e.g., Libertus et al., 2011; Starr et al., 2013b; vanMarle et al., 2014). In addition, we found that children's proficiency with manipulating ANS representations contributed additional unique variance to their symbolic math achievement that was not accounted for by ANS acuity, IQ, or executive functions. Together, these results suggest that both the acuity and manipulability of the ANS influence children's early math performance.

The majority of studies relating the ANS to symbolic math have focused on individual differences in the acuity of approximate number representations. However, the present results suggest that the manipulability of these representations

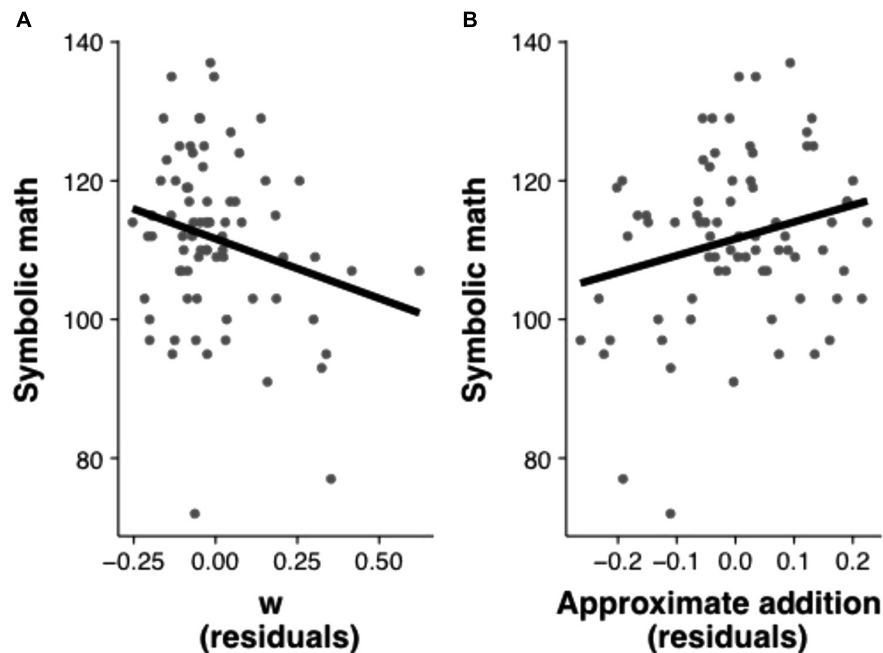


FIGURE 2 | Scatterplots illustrating the relation between w and math achievement controlling for approximate addition, IQ, and executive functions (A) and the relation between approximate addition and math achievement controlling for w , IQ, and executive functions (B).

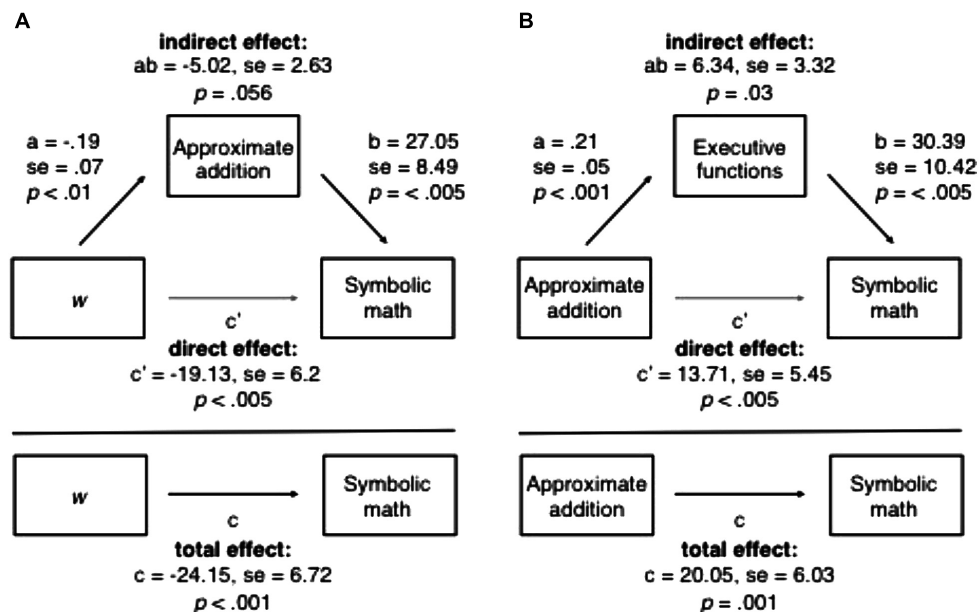


FIGURE 3 | Mediation models assessing whether (A) approximate addition mediates the relation between w and symbolic math and (B) whether executive functions mediate the relation between approximate addition and math. Note that in both cases the direct effect remains significant after accounting for the variance attributable to the mediator, indicating that the mediation is not complete.

is a second mechanism by which the ANS influences symbolic math. Although both non-symbolic numerical comparison and approximate arithmetic tasks require representing approximate numerical quantities, approximate addition additionally requires the manipulation of those quantities. Previous studies in

infants, young children, and monkeys, all of whom have no understanding of symbolic arithmetic, demonstrate that the ANS supports arithmetic operations (McCrink and Wynn, 2004; Barth et al., 2005; Cantlon and Brannon, 2007). Like symbolic arithmetic, successful approximate arithmetic requires not just

representing numerical quantities but also combining them to form summed quantity. Therefore, approximate arithmetic may provide an intuitive basis for the acquisition of symbolic arithmetic principles. Consistent with this view, we found that approximate arithmetic ability in 4.5-year-olds was a significant predictor of performance on a standardized assessment of symbolic math ability. Further, approximate arithmetic ability predicted unique variance in symbolic math scores that was not accounted for by ANS acuity, IQ, or executive functions. This result suggests that although there is a correlation between the acuity of children's ANS representations and their ability to manipulate those representations, these two factors make independent contributions to children's emerging math abilities.

Because approximate addition requires mental manipulation, it likely places a greater demand on executive functions, including working memory and updating, compared to non-symbolic numerical comparison. Given the well documented link between executive functions and math achievement in children (e.g., Bull and Scerif, 2001; St Clair-Thompson and Gathercole, 2006; Bull and Lee, 2014), one potential alternate explanation of our findings might be that the apparent link between ANS manipulability and symbolic math is actually a link between executive functions and math. However, there are multiple reasons to believe that this is not the case. First, we found that approximate addition performance was a significant predictor of math achievement even after controlling for performance on an independent executive functions task, and we found that executive functions did not mediate the relation between approximate arithmetic performance and symbolic math. Training studies in adults and children provide additional evidence that approximate arithmetic taps a cognitive skill that is separable from executive functions. These studies have found that training approximate arithmetic leads to greater improvement in symbolic arithmetic performance than does working memory training, and that approximate arithmetic training does little to improve working memory or executive functions (Park and Brannon, 2014; Park et al., 2016).

However, executive functions are a multifaceted construct (Miyake et al., 2000; Lehto et al., 2003), and we are limited in the conclusions we can draw from the use of a single executive functions task. In the present study, we used the Day/Night task (Gerstadt et al., 1994) to measure executive functions, which is similar to the task that has been used in previous studies investigating whether inhibitory control mediates the link between ANS acuity and symbolic math (Fuhs and Mcneil, 2013; Gilmore et al., 2013). This task requires both working memory (to maintain and apply the current rule) and inhibitory control (to inhibit the prepotent verbal response). However, it is possible that if we had used a separate assessment of working memory, we would have found a closer link to our approximate arithmetic task. In particular, it would be interesting to test how spatial attention interacts with approximate addition performance, given the relation between spatial attention and math achievement (Bull et al., 2008; Geary, 2011). Critically, the current results are not inconsistent with the view that executive functions contribute to successful approximate arithmetic, and disentangling the

relation between approximate arithmetic and executive functions will be an important direction for future research.

In contrast to a previous finding (Pinheiro-Chagas et al., 2014), we did not find that approximate addition performance fully mediates the relation between ANS acuity and symbolic math. Although differences in the non-symbolic comparison and approximate addition tasks used may have contributed to these inconsistent results, another possible explanation is the difference in the ages of the participants. The children in the Pinheiro-Chagas et al. (2014) study averaged 10 years of age, whereas the participants in the present study were only four. This age difference means that the children have vastly different knowledge of and experience with symbolic arithmetic. The relation between ANS acuity and symbolic math is not static with age: two recent meta-analyses have shown that the correlation between ANS acuity and symbolic math performance is strongest in young children and decreases with age (Fazio et al., 2014; Schneider et al., 2017). Therefore, it is also likely that the link between ANS manipulability and symbolic math changes with age, and this is an important area for future research.

A limitation of these data is that our approximate addition task only used numerosities between 1 and 8, which means that many of the numerosities fall within the subitizing range. However, the presence of ratio effects for both accuracy and reaction times suggests that children were not relying on subitizing to solve the addition problems. In addition, due to the speed of the addition animation, it is unlikely that children were counting the items or using a symbolic labeling strategy, and such strategies were actively discouraged. Previous work in human adults (Cordes et al., 2001; Hyde and Wood, 2011), infants (Wynn et al., 2002; Izard et al., 2008; Starr et al., 2013a), and non-human primates (Brannon and Terrace, 1998) demonstrates that the ANS can be engaged to represent both small and large numerosities. Notably, Hyde and Spelke (2011) previously suggested that stimulus complexity may predict whether small numerosities are represented by subitizing or parallel individuation versus the ANS; when stimuli are more simple, parallel individuation processes may be recruited, but when stimuli are more complex, the ANS may be recruited. This proposal can explain why infants are able to engage the ANS and succeed in discriminating two versus four elements when the displays are dynamic (Wynn et al., 2002; Starr et al., 2013a), yet fail to do so in other situations (Feigenson and Carey, 2003; Xu, 2003). The approximate addition task in the present experiment involved animated displays of moving arrays of dots, which is a situation that is likely to engage the ANS. In addition, children's approximate addition performance was ratio-dependent, meaning that accuracy was greater for trials with a 1:4 ratio compared to a 1:2 ratio. This pattern of performance, which is also seen when adults and monkeys perform approximate addition using a very similar task (Cantlon and Brannon, 2007), suggests that performance on the task is supported by the ANS. Given that approximate addition performance contributes unique variance to symbolic math achievement after controlling for ANS acuity, IQ, and executive functions, it is parsimonious to conclude that our approximate addition task is tapping

a cognitive skill not indexed by these other measures, and we believe this skill is the manipulation of approximate quantities. However, additional studies using approximate addition tasks with larger set sizes are needed to corroborate this conclusion.

CONCLUSION

The ANS endows young children with a robust sense of quantity prior to beginning formal mathematics instruction. Although many studies have provided evidence for a correlation between the fidelity of the ANS and symbolic math achievement, there remain key open questions concerning the mechanisms underlying this relation. In the present study, we provide evidence that the acuity and manipulability of the ANS have separable influences on preschoolers' early symbolic math proficiency. In particular, the influence of ANS manipulability may stem from its ability to support arithmetic operations. The shared demand for manipulating quantities may form a conceptual bridge between non-symbolic and symbolic arithmetic. Our findings therefore suggest a nuanced relation between approximate number representations and symbolic math achievement in which multiple features of the ANS contribute to the emergence of symbolic math ability in young children. In light of these results, interventions designed to target

one or both of these pathways may be differentially beneficial for children depending on their level of symbolic number knowledge and mathematical proficiency.

ETHICS STATEMENT

The study and protocol were reviewed and approved by the Duke University's Institutional Review Board. Written informed consent was obtained from the guardians of all participants and assent was obtained from all participants.

AUTHOR CONTRIBUTIONS

AS and EB conceived and planned the experiment. AS and RT collected the data. AS performed the analyses. All authors discussed the results and contributed to the final manuscript.

FUNDING

This work was funded by NSF Award 0951690 and the James McDonnell Scholar Award to EB, and an NSF GRFP and SRCD SECC Dissertation Research Funding Award to AS.

REFERENCES

- Barth, H. C., La Mont, K., Lipton, J., and Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proc. Natl. Acad. Sci.* 102, 14116–14121. doi: 10.1073/pnas.0505512102
- Barth, H. C., Lipton, J., Dehaene, S., Kanwisher, N., and Spelke, E. S. (2006). Non-symbolic arithmetic in adults and young children. *Cognition* 98, 199–222. doi: 10.1016/j.cognition.2004.09.011
- Brannon, E. M., and Terrace, H. S. (1998). Ordering of the numerosities 1 to 9 by monkeys. *Science* 282, 746–749. doi: 10.1126/science.282.5389.746
- Bull, R., Espy, K. A., and Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: longitudinal predictors of mathematical achievement at age 7 years. *Dev. Neuropsychol.* 33, 205–228. doi: 10.1080/87565640801982312
- Bull, R., and Lee, K. (2014). Executive functioning and mathematics achievement. *Child Dev. Persp.* 8, 36–41. doi: 10.1111/cdep.12059
- Bull, R., and Scerif, G. (2001). Executive functioning as a predictor of children's mathematics ability: inhibition, switching, and working memory. *Dev. Neuropsychol.* 19, 273–293. doi: 10.1207/S15326942DN1903_3
- Cantlon, J. F., and Brannon, E. M. (2007). Basic math in monkeys and college students. *PLoS Biol.* 5:e328. doi: 10.1371/journal.pbio.0050328.sg001
- Chen, Q., and Li, J. (2013). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychol.* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Cordes, S., Gelman, R., Gallistel, C. R., and Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonom. Bull. Rev.* 8, 698–707. doi: 10.3758/BF03196206
- Dehaene, S. (1997). *The Number Sense: How The Mind Creates Mathematics*. Oxford: Oxford University Press.
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Dev. Psychol.* 43, 1428–1446. doi: 10.1037/0012-1649.43.6.1428
- Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53–72. doi: 10.1016/j.jecp.2014.01.013
- Feigenson, L., and Carey, S. (2003). Tracking individuals via object-files: evidence from infants' manual search. *Dev. Sci.* 6, 568–584. doi: 10.1111/1467-7687.00313
- Fuhs, M. W., and McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: contributions of inhibitory control. *Dev. Sci.* 16, 136–148. doi: 10.1111/desc.12013
- Gallistel, C. R., and Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition* 44, 43–74. doi: 10.1016/0010-0277(92)90050-R
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Dev. Psychol.* 47, 1539–1552. doi: 10.1037/a0025510
- Gerstadt, C. L., Hong, Y. J., and Diamond, A. (1994). The relationship between cognition and action: performance of children 3 1/2–7 years old on a stroop-like day-night test. *Cognition* 53, 129–153. doi: 10.1016/0010-0277(94)90068-X
- Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., et al. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS One* 8:e67374. doi: 10.1371/journal.pone.0067374.g002
- Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2010). Non-symbolic arithmetic abilities and achievement in the first year of formal schooling in mathematics. *Cognition* 115, 394–406. doi: 10.1016/j.cognition.2010.02.002
- Ginsburg, H. P., and Baroody, A. J. (2003). *Test of Early Mathematics Ability*, 3rd Edn. Austin, TX: Pro-Ed.
- Halberda, J., and Feigenson, L. (2008). Developmental change in the acuity of the “number sense”: the approximate number system in 3-, 4-, 5-, and 6-year-olds and adults. *Dev. Psychol.* 44, 1457–1465. doi: 10.1037/a0012682
- Hubbard, E. M., Diester, I., Cantlon, J. F., Ansari, D., Van Opstal, F., and Troiani, V. (2008). The evolution of numerical cognition: from number neurons to linguistic quantifiers. *J. Neurosci.* 28, 11819–11824. doi: 10.1523/JNEUROSCI.3808-08.2008
- Hyde, D. C., Khanum, S., and Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition* 131, 92–107. doi: 10.1016/j.cognition.2013.12.007
- Hyde, D. C., and Spelke, E. S. (2011). Neural signatures of number processing in human infants: evidence for two core systems underlying numerical cognition. *Dev. Sci.* 4, 360–371. doi: 10.1111/j.1467-7687.2010.00987.x

- Hyde, D. C., and Wood, J. N. (2011). Spatial attention determines the nature of nonverbal number representation. *J. Cogn. Neurosci.* 23, 2336–2351. doi: 10.1162/jocn.2010.21581
- Izard, V., Dehaene-Lambertz, G., and Dehaene, S. (2008). Distinct cerebral pathways for object identity and number in human infants. *PLoS Biol.* 6:e11. doi: 10.1371/journal.pbio.0060011
- Kibbe, M. M., and Feigenson, L. (2015). Young children “solve for x” using the approximate number system. *Dev. Sci.* 18, 38–49. doi: 10.1111/desc.12177
- Lehto, J. E., Juujärvi, P., Kooistra, L., and Pulkkinen, L. (2003). Dimensions of executive functioning: evidence from children. *Br. J. Dev. Psychol.* 21, 59–80. doi: 10.1348/026151003321164627
- Libertus, M. E., Feigenson, L., and Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Dev. Sci.* 14, 1292–1300. doi: 10.1111/j.1467-7687.2011.01080.x
- Libertus, M. E., Feigenson, L., and Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learn. Individ. Differ.* 25, 126–133. doi: 10.1016/j.lindif.2013.02.001
- Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011). Preschoolers’ precision of the approximate number system predicts later school mathematics performance. *PLoS One* 6:e23749. doi: 10.1371/journal.pone.0023749.t003
- McCrink, K., and Wynn, K. (2004). Large-number addition and subtraction by 9-month-old infants. *Psychol. Sci.* 15, 776–781. doi: 10.1111/j.0956-7976.2004.00755.x
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734
- Park, J., Bermudez, V., and Roberts, R. C. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *J. Exp.* 152, 278–293. doi: 10.1016/j.jecp.2016.07.011
- Park, J., and Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychol. Sci.* 24, 2013–2019. doi: 10.1177/0956797613482944
- Park, J., and Brannon, E. M. (2014). Improving arithmetic performance with number sense training: an investigation of underlying mechanism. *Cognition* 133, 188–200. doi: 10.1016/j.cognition.2014.06.011
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012
- Pinheiro-Chagas, P., Wood, G., Knops, A., Krinzinger, H., Lonnemann, J., Starling-Alves, I., et al. (2014). In how many ways is the approximate number system associated with exact calculation? *PLoS One* 9:e111155. doi: 10.1371/journal.pone.0111155.s002
- Reynolds, C. R., and Kamphaus, R. W. (2003). *Reynolds Intellectual Assessment Scales*. Lutz, FL: Psychological Assessment Resources, Inc.
- Rosseel, Y. (2012). *Lavaan: An R Package for Structural Equation Modeling and More. Version 0.5–12 (BETA)*
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372
- St Clair-Thompson, H. L., and Gathercole, S. E. (2006). Executive functions and achievements in school: shifting, updating, inhibition, and working memory. *Q. J. Exp. Psychol.* 59, 745–759. doi: 10.1080/17470210500162854
- Starr, A., Libertus, M. E., and Brannon, E. M. (2013a). Infants show ratio-dependent number discrimination regardless of set size. *Infancy* 18, 927–941. doi: 10.1111/inf.12008
- Starr, A., Libertus, M. E., and Brannon, E. M. (2013b). Number sense in infancy predicts mathematical abilities in childhood. *Proc. Natl. Acad. Sci.* 110, 18116–18120. doi: 10.1073/pnas.1302751110/-DCSupplemental/sm01.avi
- vanMarle, K., Chu, F. W., Li, Y., and Geary, D. C. (2014). Acuity of the approximate number system and preschoolers’ quantitative development. *Dev. Sci.* 17, 492–505. doi: 10.1111/desc.12143
- Wynn, K., Bloom, P., and Chiang, W.-C. C. (2002). Enumeration of collective entities by 5-month-old infants. *Cognition* 83, B55–B62. doi: 10.1016/S0010-0277(02)00008-2
- Xu, F. (2003). Numerosity discrimination in infants: evidence for two systems of representations. *Cognition* 89, B15–B25. doi: 10.1016/S0010-0277(03)00050-7

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Starr, Tomlinson and Brannon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



A Mechanistic Study of the Association Between Symbolic Approximate Arithmetic Performance and Basic Number Magnitude Processing Based on Task Difficulty

Wei Wei^{1*}, Wanying Deng¹, Chen Chen¹, Jie He¹, Jike Qin² and Yulia Kovas^{3,4}

¹ Department of Psychology and Behavioral Sciences, Zhejiang University, Hangzhou, China, ² Department of Psychology, The Ohio State University, Columbus, OH, United States, ³ Department of Psychology, Goldsmiths, University of London, London, United Kingdom, ⁴ Laboratory for Cognitive Investigations and Behavioural Genetics, Tomsk State University, Tomsk, Russia

OPEN ACCESS

Edited by:

Jingguang Li,
Dali University, China

Reviewed by:

Jiwei Si,
Shandong Normal University, China
Mathieu Guillaume,
University of Luxembourg,
Luxembourg

*Correspondence:

Wei Wei
weiwei820@zju.edu.cn

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 24 April 2018

Accepted: 06 August 2018

Published: 11 September 2018

Citation:

Wei W, Deng W, Chen C, He J, Qin J
and Kovas Y (2018) A Mechanistic
Study of the Association Between
Symbolic Approximate Arithmetic
Performance and Basic Number
Magnitude Processing Based on Task
Difficulty. *Front. Psychol.* 9:1551.
doi: 10.3389/fpsyg.2018.01551

Two types of number magnitude processing – semantic and spatial – are significantly correlated with children's arithmetic performance. However, it remains unclear whether these abilities are independent predictors of symbolic approximate arithmetic performance. The current study addressed this question by assessing 86 kindergartners (mean age of 5 years and 7 months) on semantic number processing (number comparison task), spatial number processing (number line estimation task), and symbolic approximate arithmetic performance with different levels of difficulty. The results showed that performance on both tasks of number magnitude processing was significantly correlated with symbolic approximate arithmetic performance, but the strength of these correlations was moderated by the difficulty level of the arithmetic task. The simple symbolic approximate arithmetic task was equally related to both tasks. In contrast, for more difficult symbolic approximate arithmetic tasks, the contribution of number comparison ability was smaller than that of the number line estimation ability. These results indicate that the strength of contribution of the different types of numerical processing depends on the difficulty of the symbolic approximate arithmetic task.

Keywords: symbolic approximate arithmetic, kindergartner, number processing, number line estimation, number comparison, task difficulty

INTRODUCTION

Arithmetic competency is an important aspect of mathematical ability. Over the past few decades, many studies have investigated the cognitive mechanisms underlying exact arithmetic ability (De Smedt et al., 2013; Moeller et al., 2015; see Arsalidou and Taylor, 2011; Schneider et al., 2017, for reviews). However, less is known about the cognitive mechanisms underlying symbolic approximate arithmetic calculations, such as solving the following task: “give an approximate answer for 38×21 in 5 s.”

Symbolic approximate arithmetic performance refers to the ability to provide an approximate answer rather than an exact one (Gilmore et al., 2007; McNeil et al., 2011; Xenidou-Dervou et al., 2015). This ability plays an important role in mathematical learning (Xenidou-Dervou et al., 2013). This importance has begun to receive recognition by educational authorities. For example, symbolic approximate arithmetic performance

is listed as an important part of mathematical learning by The National Council of Supervisors of Mathematics and the National Council of Teachers of Mathematics in the United States (1989), as well as by the Ministry of Education in Japan (1989). Understanding cognitive mechanisms underlying symbolic approximate arithmetic performance will help with the designing of curricula that develop symbolic approximate arithmetic skills.

Recent research has begun to provide insights into these mechanisms. Research has suggested that, unlike exact arithmetic ability, symbolic approximate arithmetic performance may not be influenced by culture (Reys and Yang, 1998), language, or education (Spelke and Tsivkin, 2001; Nys et al., 2013). For example, recent research has found that preschool children can solve symbolic approximate arithmetic problems with large numbers, even if they cannot provide exact answers (Gilmore et al., 2007; McNeil et al., 2011; Xenidou-Dervou et al., 2015).

Two types of tasks have typically been used to assess basic numerical magnitude processing: *the number magnitude comparison task*, primarily tapping into number semantic processing (Pinel et al., 2001; Rousselle and Noël, 2007; Holloway and Ansari, 2009; see De Smedt et al., 2013, for a review) and *the number line estimation task*, primarily tapping into spatial number processing (Dehaene et al., 2003; Hubbard et al., 2005; Siegler and Ramani, 2008; Berteletti et al., 2012; see Moeller et al., 2015, for a review). These abilities are commonly referred to in the literature as *number sense*, although recent research suggests that it is a highly heterogeneous concept (e.g., Berch, 2005; Halberda et al., 2008; Tosto et al., 2017; see Cohen Kadosh et al., 2008; De Smedt et al., 2013, for reviews).

Number semantic processing and number spatial processing are both correlated with exact arithmetic processing. For example, correlations between exact arithmetic processing and number semantic processing have been found in typically developing children (Durand et al., 2005; Bartelet et al., 2014; Vanbinst et al., 2015), in children with developmental dyscalculia (Landerl et al., 2004; Mussolin et al., 2010), as well as in training studies (Wilson et al., 2006, 2009). Similarly, correlations between exact arithmetic processing and number spatial processing have been shown in typically developing children (Booth and Siegler, 2008; Laski and Yu, 2014), as well as in training studies (Siegler and Ramani, 2008; Kucian et al., 2011).

It is possible that number semantic processing and number spatial processing may relate to arithmetic abilities through a common mechanism. For example, Laski and Siegler (2007) examined the performance on number line estimation and number comparison tasks in 5–8-year-old children and observed strong associations between the two tasks within each grade. However, other studies suggest that the two abilities influence arithmetic performance through different mechanisms, as the two are at least partially independent. For example, Sasanguie and Reynvoet (2013) found that children in grades 1–3 who were faster at comparing numbers performed better on a timed arithmetic test 1 year later. In contrast, no significant associations were found between performance on symbolic number line estimation task and a timed arithmetic test. Recent data provided by Linsen et al. (2014) further showed significant associations

between number processing (including number comparison and number line estimation tasks) and the more specific mathematical skill of mental subtraction. In their study, the association between number comparison and mental subtraction remained after controlling for the number line estimation, whereas the association between number line estimation and mental subtraction disappeared after controlling for the number comparison task.

Both number magnitude comparison ability (Gilmore et al., 2007) and number line estimation ability (Gunderson et al., 2012) have been found to be associated with children's performance on symbolic approximate arithmetic tasks. However, most previous studies have examined only one of these basic numerical processing tasks at a time, which makes it difficult to evaluate the extent to which they differentially predict symbolic approximate arithmetic performance.

It is necessary to involve two number magnitude processing tasks in one study to investigate their differential influence on symbolic approximate arithmetic performance. We put forward our first hypothesis, "*Number semantic processing and spatial processing are significantly correlated with the performance of symbolic approximate arithmetic tasks.*"

It is possible that relations between number magnitude processing and arithmetic processing are moderated by the difficulty of the arithmetic task, in that different difficulty levels of arithmetic problems rely on semantic and spatial number tasks to a different extent. For example, a correlation between performance on exact arithmetic processing and number semantic processing has been observed in simple exact arithmetic tasks (e.g., single-digit arithmetic problems) (Landerl et al., 2004; Durand et al., 2005; Bugden et al., 2012; Bartelet et al., 2014; Vanbinst et al., 2015). In contrast, other studies found a correlation between exact arithmetic processing and number spatial processing, which has been observed in difficult arithmetic tasks, such as two-digit or three-digit arithmetic problems. Complex mathematical problems are more dependent on spatial processing when compared with simple problems. For the simple arithmetic problems, participants retrieved the answers from long-term semantic working memory (Geary et al., 1996; LeFevre et al., 1996; Delazer and Benke, 1997; McLean and Hitch, 1999), whereas much more visuospatial processing was involved in the processing of complex arithmetic problems (Zago et al., 2001; Berteletti et al., 2015).

It is unclear if the symbolic approximate arithmetic task has the same effects as the exact arithmetic task. We put forward our second hypothesis, "*The relations between number semantic processing and spatial processing and symbolic approximate arithmetic ability vary as a function of the difficulty of the symbolic approximate arithmetic tasks.*"

MATERIALS AND METHODS

Participants

A total of 94 typically developing children from middle-to-high socioeconomic status (SES) backgrounds were recruited from three kindergartens in the urban area of Hangzhou, China. Data

of eight children were removed from the analyses because they either correctly answered at least one question in the probe stage (see Procedure for details) of the symbolic approximate arithmetic task ($n = 5$) or did not complete all the tasks ($n = 3$). The final sample included 86 children (45 boys and 41 girls). Their mean age was 5 years and 7 months (ranging from 5 years and 1 month to 6 years and 3 months). Similar to most countries, formal mathematics education starts in the first year of elementary school in China; therefore, children in this study were assessed prior to receiving formal mathematical instruction. Permission to conduct the study was given by the principals of the kindergartens. Written informed consents were obtained from all the parents. The study was approved by the principals of the kindergartens and the ethics committee at the Zhejiang University, China.

Measures

Symbolic Approximate Arithmetic Task

The symbolic approximate arithmetic task was adapted from Gilmore et al. (2007). Arithmetical questions were presented both visually on a computer screen and verbally by the experimenter. The children had to indicate which side of the screen had a larger numerical magnitude through mental arithmetic. For example, on the screen, one cartoon character first received a bag of candies marked with the number 13 and then received a second bag marked with the number 22. Another cartoon character received a bag of candies marked with the number 28. The children needed to determine which character had more candies in total. Each trial would remain on the screen until the participants responded. This task consisted of 5 practice problems and 24 formal problems. The formal problems were divided into three levels of difficulty according to the ratios of the sum of the problem to the comparison number, that is, 4:7 (Level 1, the easiest level), 4:6 (Level 2, the medium level), and 4:5 (Level 3, the hardest level). The numbers ranged from 6 to 56. The exact answer was larger than the comparison number in half of the problems, whereas it was smaller in the other half of the problems. The formal problems part was split into three blocks. Each block included eight problems. The difficulty levels varied within each block. Error rate was used as the index of performance. There was no time limitation for the children's responses. Spearman-Brown corrected split half reliability was $r = 0.76$.

Symbolic Number Comparison Task

The symbolic number comparison task was adapted from Gilmore et al.'s (2007) study. In the number comparison task, two two-digit numbers were used. The numbers were presented on a computer screen at the same time, and the children were asked to judge which number was larger. A total of 5 practice problems were followed by 24 formal problems. The children were required to make the judgment. If the children chose the left number, they pressed the "F" key on the computer keyboard; if they chose the right number, they pressed the "J" key. Each trial would remain on the screen until the participants responded. The order of the presentation was random for each participant.

After each practice trial, the children would see a smiling face on the screen if they responded correctly or a crying face if they responded incorrectly. Only children with accuracy above 60% in the practice problems would be given the formal problems. No feedback was given following the formal trials. The index of performance was the error rate. Spearman-Brown corrected split half reliability was $r = 0.83$.

Number Line Estimation Task

This task was adopted from Siegler and Opfer's (2003) study. The children were given 28 sheets of paper, 2 for practice trials and 26 for formal trials, each with the same 25 cm number line printed in the center and a number between 0 and 100 printed 2 cm above the middle of the line. The experimenter initially told the children the following: "Each number has its own specific position on the number line and you should mark the position where you think the number actually is on the line using a pencil. Try your best to do it exactly." For two practice problems, the children were asked to mark the location of the number 50. If they failed, the experimenter would help them to find the correct location. The formal problems had "0" written below the start of the number line and "100" written below the end point. A total of 26 trials were held, respectively, for the 26 numbers to be estimated. The numbers used in the experiment (3, 4, 6, 8, 12, 14, 17, 18, 21, 24, 25, 29, 33, 39, 42, 48, 52, 57, 61, 64, 72, 79, 81, 84, 90, and 96) were taken from Booth and Siegler's (2006) study. The order was presented randomly for each child. The main performance index was the percent of absolute error [$PAE = (|estimate - estimated\ quantity| / \text{scale of estimates}) \times 100$], where estimate is the participant's answer, estimated quantity is the correct answer, scale of estimates is 100 in the current study. PAE reflects the accuracy of numerical estimation and has been used in a large number of studies (Booth and Siegler, 2008; Laski and Yu, 2014; Xenidou-Dervou et al., 2015). A smaller PAE indicates more accurate numerical estimation. Spearman-Brown corrected split half reliability was $r = 0.81$.

Procedure

The symbolic approximate arithmetic and number comparison tasks were presented on a laptop with a 15-inch monitor. The stimuli for the symbolic approximate arithmetic and number comparison tasks were presented using Presentation® software (version 0.71; Neurobehavioral Systems, Berkeley, CA, United States). The number line estimation was a paper-and-pencil task.

For all experimental measures, the children were tested one by one in a quiet room in the kindergarten, accompanied by an experimenter. The children performed the tasks in the following order: the symbolic number comparison task, the number line estimation task, and the approximate addition arithmetic task. A short break of about 2 min was provided between each task.

In order to prevent the children from performing exact calculations in the symbolic approximate arithmetic task, a probe stage was then conducted. In the probe stage, the children were asked to provide the exact answers for two problems, which were

chosen randomly from the ones they had performed correctly in the formal part of the symbolic approximate arithmetic task. The data of the children who correctly answered at least one question in the probe stage were removed from the analysis. This approach ensured that the participants were unable to perform the exact calculations.

The whole test took approximately 25 min for each child. Following the experiment, each child received a sticker as a reward.

Data Analysis

The following data analysis was performed using the SPSS 19.0 software (SPSS Inc., Chicago, IL, United States). Analyses were performed on error rate for the symbolic approximate arithmetic and symbolic number comparison tasks and on PAE for the number line estimation task. No participants were outliers (three *SD* above or below the group mean) for each task. Error rates and PAE have the same direction. First, we calculated the correlation coefficients between the symbolic approximate arithmetic task and the two basic number processing tasks after controlling for gender and age. To explore the specific roles of the number comparison and number line estimation abilities in different levels of symbolic approximate arithmetic performance, we conducted hierarchical regression analyses.

RESULTS

Descriptive and Preliminary Analysis

All dependent measures and predictors are presented in **Table 1**. The correlations between the different levels of symbolic approximate arithmetic task and basic numerical magnitude processing task are presented in **Table 2** (controlling for gender and age). First, a series of analyses were conducted in order to verify that our children were able to perform the tasks. The error rate of approximate addition showed that children performed below chance level (50%) on all three levels (Level 1: $M = 28\%$, $t_{85} = -11.060$, $p < 0.001$; Level 2: $M = 33\%$, $t_{85} = -7.421$, $p < 0.001$; and Level 3: $M = 38\%$, $t_{85} = -5.881$, $p < 0.001$). These results were similar to Gilmore et al.'s (2007) study, which had a 26.7% error rate for approximate addition problems. The children's error rate for number comparison tasks was also below 50% ($M = 22\%$, $t_{85} = -13.036$, $p < 0.001$), which was similar to Gilmore et al.'s (2007) study (19.6% error rate for number comparison tasks). The children's mean PAE was 20.65%, which was similar to the previous studies (Booth and Siegler, 2006, $M = 24\%$).

In addition, a repeated measure ANOVA was conducted to test the effect of difficulty on symbolic approximate arithmetic performance, $F_{(2,170)} = 13.263$, $p < 0.001$. The *post hoc* results showed that as the difficulty increased, the accuracy of symbolic approximate arithmetic performance decreased. Error rate for Level 1 of symbolic approximate arithmetic performance was significantly lower than that for Level 3 [$F_{(1,85)} = 24.638$, $p < 0.001$], Level 1 was significantly lower than Level 2 [$F_{(1,85)} = 9.328$, $p = 0.003$], and Level 2 was significantly lower than Level 3 [$F_{(1,85)} = 5.012$, $p = 0.028$].

TABLE 1 | Descriptive statistics of kindergartners' performance on measures of symbolic approximate arithmetic ability, symbolic number comparison ability, and number line estimation ability.

Task	Index	Mean (SD)	Range
Symbolic approximate arithmetic	Error rate	0.33 (0.16)	0.00 ~ 0.58
Level 1 (4:7)	Error rate	0.28 (0.19)	0.00 ~ 0.62
Level 2 (4:6)	Error rate	0.33 (0.21)	0.00 ~ 0.75
Level 3 (4:5)	Error rate	0.38 (0.19)	0.00 ~ 0.88
Symbolic number comparison	Error rate	0.22 (0.20)	0.00 ~ 0.62
Number line estimation	PAE	20.65% (8.55)	3.68% ~ 41.16%

PAE = percent of absolute error.

TABLE 2 | Correlations between the basic numerical magnitude processing tasks and different difficulty levels of the symbolic approximate arithmetic tasks after controlling for gender and age.

	1	2	3	4	5
Symbolic approximate arithmetic					
(1) Level 1	—				
(2) Level 2	0.59***	—			
(3) Level 3	0.42***	0.48***	—		
(4) All Levels	0.81***	0.86***	0.77***	—	
Basic numerical processing					
(5) Number comparison	0.33**	0.40***	0.35**	0.44***	—
(6) Number line estimation	0.29**	0.39***	0.39***	0.44***	0.44***

** $p < 0.01$ and *** $p < 0.001$. "All Levels" was the composite score of Level 1, Level 2, and Level 3.

Hierarchical Regression Analysis

Two models of hierarchical regression analysis were carried out to further examine the relationships among number comparison performance, number line estimation performance, and different levels of symbolic approximate arithmetic performance. The error rates of the three different levels of symbolic approximate arithmetic performance were the outcome variables.

The first regression model tested whether number magnitude comparison ability was associated with different levels of symbolic approximate arithmetic performance after controlling for gender, age, and number line estimation performance. Gender and age were entered into the model first, following which the PAE of number line estimation performance and the error rate of number comparison performance were entered, respectively. For Level 1 and Level 2, number comparison performance was a significant predictor of symbolic approximate arithmetic performance after controlling for gender, age, and number line estimation performance. However, for the most difficult level (Level 3), number comparison performance was not a significant predictor. For Level 1, number line estimation performance was not a significant predictor for symbolic approximate arithmetic performance after number comparison performance was entered into Model 1. However, for Levels 2 and 3, number line estimation performance continued to be a significant predictor of symbolic approximate arithmetic performance, even when number comparison performance was entered into the model.

The *R*-square change carried by number comparison performance decreased, becoming 4.2% for Level 1, 4.9% for Level 2, and 3.0% for Level 3, after controlling for number line estimation performance, gender, and age. For difficulty Levels 1 and 2, number comparison performance significantly improved the fit of the model, whereas for Level 3, it did not significantly improve the fit of the model (Table 3, Model 1).

To assess the relative contribution of performance on the two tasks to the 3 levels of difficulty on symbolic approximate arithmetic performance, a second regression model was conducted, reversing the order of entry. Gender and age were entered first, followed by the error rate of number comparison performance and the PAE of number line estimation performance, respectively. For Level 1, number line estimation performance was not a significant predictor of symbolic approximate arithmetic performance after controlling for gender, age, and number comparison performance. For Level 2, the regression coefficient of both number line estimation performance and number comparison performance were significant. For the most difficult level (Level 3), number comparison performance was not a significant predictor after number line estimation performance was entered into the model. For Levels 1 and 2, number comparison performance was still a significant predictor for symbolic approximate arithmetic performance after number line estimation performance was entered into Model 2. However, for Level 3, number comparison performance was not a significant predictor of symbolic approximate arithmetic performance when the number line estimation performance was entered into the model. As symbolic approximate arithmetic performance became more difficult, the *R*-square change uniquely carried by number line estimation performance increased gradually, becoming 2.1% for Level 1, 4.7% for Level 2, and 5.9% for Level 3 after number comparison was controlled. It should be noted that number line estimation performance significantly improved the fit of the model for Levels 2 and 3 of symbolic approximate arithmetic performance (Table 3, Model 2).

DISCUSSION

The current study aimed to investigate the relations between two basic numerical magnitude processing abilities (semantic and spatial) and symbolic approximate arithmetic performance. The results supported the two hypotheses proposed. First, both number magnitude comparison and number line estimation abilities were significantly correlated with the performance on symbolic approximate arithmetic tasks. Second, the relations between the two basic numerical magnitude processing abilities and symbolic approximate arithmetic performance varied with a change in the difficulty of the symbolic approximate arithmetic tasks; with an increase in the difficulty of the symbolic approximate arithmetic task, the contribution of number magnitude comparison ability decreased, whereas the contribution of number line estimation ability increased. The results indicate that number line estimation ability plays a particularly important role in symbolic approximate arithmetic performance with a higher level of difficulty.

Similarity Between Semantic and Spatial Number Magnitude Processing Abilities

Previous studies have found significant relations between arithmetic ability and number magnitude comparison or number line estimation abilities (Booth and Siegler, 2008; Gunderson et al., 2012; Sasanguie and Reynvoet, 2013; Bartelet et al., 2014), as well as significant correlations between number magnitude comparison ability and number line estimation ability (Laski and Siegler, 2007). In our study, we found that both numerical magnitude tasks correlated with each other and had significant correlations with symbolic approximate arithmetic performance, which was consistent with previous studies (Gillmore et al., 2007; Laski and Siegler, 2007; Gunderson et al., 2012).

Performance on number magnitude comparison and number line estimation tasks may rely on the same underlying representation, similar to a compressed mental number line (Gallistel and Gelman, 1992; Dehaene, 2011). Specifically, a mental number line representation implies that magnitudes are represented as a Gaussian distribution around the true location of each specific number, with partially overlapping representations for nearby numbers. Such a representational organization leads to greater difficulty in discriminating between nearby numbers. This is reflected in both higher error rates and longer reaction times for near distance pairs when compared with far distance ones in a comparison task (distance effect) and in the inaccurate estimation of the location of specific numbers in a number line task within the range of familiar numbers. Because of the common representation, symbolic approximate arithmetic performance is significantly correlated with both basic numerical magnitude processing tasks. And both basic numerical magnitude processing tasks were related to each other.

Differences Between Semantic and Spatial Number Magnitude Processing Abilities

The results of the current study were consistent with previous studies that demonstrated that number comparison and number line estimation abilities play different roles in arithmetic performance with different levels of difficulty (Sasanguie and Reynvoet, 2013; Linsen et al., 2014).

The different contributions of the two basic numerical magnitude processes to symbolic approximate arithmetic performance could be explained by the evidence provided for the dissociation between number comparison and number line estimation abilities. For example, a patient who had damage to the left posterior parietal lobe was impaired in the ability to process the relative positions of numbers, while the ability to perform tasks that required the processing of the meaning of numerical magnitude was preserved (Turconi and Seron, 2002). In addition, functional magnetic resonance imaging (fMRI) studies and event-related potential (ERP) studies showed separate neural circuits or brain signatures for processing numerical magnitude information and numerical spatial information. Researchers have found the different spatial and temporal courses between numerical processing and ordinal processing using ERPs (Turconi et al., 2004; Rubinsten et al.,

TABLE 3 | Hierarchical regression analysis predicting performance on three different levels of symbolic approximate arithmetic ability.

	Predictors	Model 1											
		Level 1				Level 2				Level 3			
		B (SE)	β	R ²	ΔR ²	B (SE)	β	R ²	ΔR ²	B (SE)	β	R ²	ΔR ²
Step 1	Gender	−0.05 (0.04)	−0.12			−0.03 (0.04)	−0.08			−0.06 (0.04)	−0.16		
	Age	−0.15 (0.06)*	−0.25*	8.4%	8.4%	−0.20 (0.07)**	−0.30**	10.2%	10.2%	−0.15 (0.06)*	−0.24*	8.8%	8.8%
Step 2	Gender	−0.01 (0.04)	−0.03			0.02 (0.04)	0.05			−0.01 (0.04)	−0.03		
	Age	−0.13 (0.06)*	−0.21*			−0.16 (0.07)*	−0.24*			−0.11 (0.06)	−0.18		
Step 3	Number line estimation	0.01 (0.002)**	0.30**	16.1%	7.7%	0.01 (0.003)***	0.40***	24.0%	13.8%	0.01 (0.002)***	0.40***	22.9%	14.1%
	Gender	−0.02 (0.04)	−0.06			0.003 (0.04)	0.008			−0.02 (0.04)	−0.06		
Model 2													
	Age	−0.09 (0.06)	−0.14			−0.11 (0.07)	−0.17			−0.08 (0.06)	−0.12		
	Number line estimation	0.004 (0.003)	0.18			0.01 (0.003)*	0.27*			0.01 (0.003)*	0.30*		
Step 1	Number comparison	0.23 (0.11)*	0.25*	20.3%	4.2%	0.28 (0.12)*	0.27*	28.9%	4.9%	0.20 (0.11)	0.21	26.0%	3%
	Gender	−0.05 (0.04)	−0.12			−0.03 (0.04)	−0.08			−0.06 (0.04)	−0.16		
Step 2	Age	−0.15 (0.06)*	−0.25*	8.4%	8.4%	−0.20 (0.07)**	−0.30**	10.2%	10.2%	−0.15 (0.06)*	−0.24*	8.8%	8.8%
	Gender	−0.04 (0.04)	−0.12			−0.03 (0.04)	−0.08			−0.06 (0.04)	−0.15		
Step 3	Age	−0.09 (0.06)	−0.14			−0.11 (0.07)	−0.17			−0.07 (0.06)	−0.12		
	Number comparison	0.32 (0.10)**	0.33**	18.2%	9.8%	0.42 (0.11)****	0.40***	24.2%	14.0%	0.34 (0.10)**	0.36**	20.0%	11.2%
	Gender	−0.02 (0.04)	−0.06			0.003 (0.04)	0.008			−0.02 (0.04)	−0.06		
	Age	−0.09 (0.06)	−0.14			−0.11 (0.07)	−0.17			−0.08 (0.06)	−0.12		
	Number comparison	0.23 (0.11)*	0.25*			0.28 (0.12)*	0.27*	28.9%	4.7%	0.20 (0.11)	0.21	26.0%	5.9%
	Number line estimation	0.004 (0.003)	0.18	20.3%	2.1%	0.01 (0.003)*	0.27*	28.9%		0.01 (0.003)*	0.30*	26.0%	

* $p < 0.05$; ** $p < 0.01$; and *** $p < 0.001$.

2013). Using fMRI, researchers have also found that the ordinal processing and cardinal number processing have a separate brain activation in the intraparietal sulcus (Tang et al., 2008). Furthermore, a behavioral study failed to find a transfer effect between number comparison and number line estimation abilities (Maertens et al., 2016).

Difficulty of Symbolic Approximate Arithmetic Performance and Number Magnitude Processing

The current study found that when the difficulty of symbolic approximate arithmetic tasks increased, the number line estimation ability contributed more to the symbolic approximate arithmetic performance, whereas the number comparison ability contributed less.

One possible explanation is that the numerical magnitude comparison ability develops earlier than the number line estimation ability. This hypothesis is supported by a variety of findings. First, studies in developmental psychology have shown that the ability to process quantities is part of a “cognitive core knowledge,” recent studies have found that accuracy on a symbolic number comparison task in the range of 1–100 reaches about 90% in 6-year-old kindergartners (Kolkman et al., 2013). However, their performance on number line estimation tasks in the same numerical range continues to show sustained development across grades 1–3 (Booth and Siegler, 2006). Second, a developmental model of number acquisition (Von Aster and Shalev, 2007) has described the development of numerical cognition in four steps with the learning of the basic meaning of numbers as the first step, the verbal learning of number words as the second step, the connection between the Arabic number system and the former two steps as the third step, and the numerical spatial representation that develops during the school period as a result of the development of the first three steps as the fourth step. Altogether, this evidence suggests that the numerical comparison ability develops earlier than the number line estimation ability. According to the cognitive load theory (CLT), if the extraneous cognitive load (corresponding to symbolic approximate arithmetic performance in our study) was not high, automated schema in long-term memory (corresponding to number semantic processing ability) would be used to solve the problem; whereas if the extraneous cognitive load was high, a complex schema (corresponding to number spatial processing ability) should be developed to solve the complex problems (Sweller et al., 1998; Van Merriënboer and Sweller, 2005). Because number semantic processing ability develops earlier than number spatial processing ability, number semantic processing ability should develop earlier in the cognitive process. When the symbolic approximate arithmetic task was simple (the extraneous cognitive load is low), the children's number comparison ability (automated schema) was used first, whereas when the symbolic approximate arithmetic task became more difficult, the number line estimation ability (developed schema) gradually began to operate.

The second possible explanation is that complex mathematical problems depend on spatial processing ability when compared

with simple problems. Behavioral and neuroimaging studies have found that as the difficulty of the mathematical problem increases, spatial ability plays a more significant role. A developmental study (Sasanguie et al., 2012) found that when mathematical ability was tested with complex problems, the number line estimation ability predicted performance more strongly than the number comparison ability. Previous neuroimaging studies for children have demonstrated that complex arithmetic problem activates the parietal lobe more than simple arithmetic problem (Menon et al., 2000; De Smedt et al., 2011; Ashkenazi et al., 2012; Berteletti et al., 2015). Moreover, the number line estimation ability has significant correlations with complex arithmetic performance in the brain. One neuroimaging study showed that number line estimation ability was related to arithmetic performance by comparing the activation of the parietal lobe for simple and complex arithmetic problems (Berteletti et al., 2015). A training study showed that less activation occurred in the parietal lobe in response to a number task following number line estimation training (Kucian et al., 2011). Spatial information always depends on the parietal lobe (see Zacks, 2008, for a meta-analysis and review). Spatial attention and visuospatial working memory abilities were needed to solve the complex arithmetic problems when compared with the simple ones (Zago et al., 2001; Zago and Tzourio-Mazoyer, 2002; Berteletti et al., 2015). In this study, with the increase in difficulty, the sum of the problem was closer to the comparison number. Participants had to rely on much more spatial attention and visuospatial working memory process to retrieve the approximate answers from the mental number line (Knops et al., 2009).

A limitation of the present study is given by the consideration of the SES in which this sample reflected. Specifically, recent findings indicated that SES backgrounds can affect children's performance on the symbolic approximate arithmetic tasks, suggesting that those from middle-to-high SES backgrounds performed significantly better than age-matched peers from low SES backgrounds (McNeil et al., 2011). Our data were all collected from kindergartens in urban areas, which were assumed to be representatives of middle-to-high SES backgrounds, thereby the question of whether this result will also generalize to other samples remains to be further investigated.

The current study provides new insights into the cognitive mechanisms of symbolic approximate arithmetic performance for kindergartners. The finding that symbolic approximate arithmetic ability is related to basic numerical magnitude processing implies that performance on symbolic approximate arithmetic tasks may be improved through basic number magnitude processing training. The results also suggest that for complex arithmetic tasks, number spatial ability may be more essential. Training studies have found that spatial representation of numbers could be taught using games (Siegler and Ramani, 2008; Kucian et al., 2011; De Smedt et al., 2013). These studies used computer games or board games to teach the spatial presentation of numbers. Feedback, provided in the game guides, helps children to learn the correct position of numbers. Future studies should be conducted to explore the effects of such game

training on different levels of arithmetic performance, as well as on exact arithmetic ability in the same study.

The participants in this study were Chinese kindergartners. Previous cross-cultural studies had found that Chinese kindergartners had superiority in exact arithmetic ability (Rodic et al., 2015) and number line estimation ability (Siegler and Mu, 2008). This superiority could be because of the base-10 structure system of number name which could help Chinese kindergartners to count and understand the meaning of numbers (Miller et al., 1995). Secondly, Chinese children have more information related to numbers in daily life (Kelly et al., 1999). For example, Chinese people use numbers to name months, that is, January in Chinese is “the first month,” February is “the second month,” and so on. Chinese parents typically have higher expectations regarding mathematical achievement when compared with western parents, which influences the Chinese parents to teach their children mathematics at home before entering primary school (see Ng and Rao, 2010, for a review). Further studies could be carried

out to investigate if the cultural differences would influence the performance of symbolic approximate arithmetic tasks.

AUTHOR CONTRIBUTIONS

WW designed the experiments and drafted the manuscript. WD collected and analyzed the data and drafted the manuscript. CC collected the data. JH and JQ provided methodological advice. YK revised the manuscript.

FUNDING

This research was supported by the National Natural Science Foundation of China (No. 31500902) and by the Zhejiang Provincial Natural Science Foundation of China (No. LQ15C090001).

REFERENCES

- Arsalidou, M., and Taylor, M. J. (2011). Is $2+2=4$? Meta-analyses of brain areas needed for numbers and calculations. *Neuroimage* 54, 2382–2393. doi: 10.1016/j.neuroimage.2010.10.009
- Ashkenazi, S., Rosenberg-Lee, M., Tenison, C., and Menon, V. (2012). Weak task-related modulation and stimulus representations during arithmetic problem solving in children with developmental dyscalculia. *Dev. Cogn. Neurosci.* 2, S152–S166. doi: 10.1016/j.dcn.2011.09.006
- Bartelet, D., Vaessen, A., Blomert, L., and Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *J. Exp. Child Psychol.* 117, 12–28. doi: 10.1016/j.jecp.2013.08.010
- Berch, D. B. (2005). Making sense of number sense: implications for children with mathematical disabilities. *J. Learn. Disabil.* 38, 333–339. doi: 10.1177/00222194050380040901
- Berteletti, I., Lucangeli, D., and Zorzi, M. (2012). Representation of numerical and non-numerical order in children. *Cognition* 124, 304–313. doi: 10.1016/j.cognition.2012.05.015
- Berteletti, I., Man, G., and Booth, J. R. (2015). How number line estimation skills relate to neural activations in single digit subtraction problems. *Neuroimage* 107, 198–206. doi: 10.1016/j.neuroimage.2014.12.011
- Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42, 189–201. doi: 10.1037/0012-1649.41.6.189
- Booth, J. L., and Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Dev.* 79, 1016–1031. doi: 10.1111/j.1467-8624.2008.01173.x
- Bugden, S., Price, G. R., McLean, D. A., and Ansari, D. (2012). The role of the left intraparietal sulcus in the relationship between symbolic number processing and children's arithmetic competence. *Dev. Cogn. Neurosci.* 2, 448–457. doi: 10.1016/j.dcn.2012.04.001
- Cohen Kadosh, R., Lammertyn, J., and Izard, V. (2008). Are numbers special? *Prog. Neurobiol.* 84, 132–147.
- De Smedt, B., Holloway, I. D., and Ansari, D. (2011). Effects of problem size and arithmetic operation on brain activation during calculation in children with varying levels of arithmetic fluency. *Neuroimage* 57, 771–781. doi: 10.1016/j.neuroimage.2010.12.037
- De Smedt, B., Noël, M.-P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends Neurosci. Educ.* 2, 48–55. doi: 10.1016/j.tine.2013.06.001
- Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics*. Oxford: OUP.
- Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239
- Delazer, M., and Benke, T. (1997). Arithmetic facts without meaning. *Cortex* 33, 697–710. doi: 10.1016/S0010-9452(08)70727-5
- Durand, M., Hulme, C., Larkin, R., and Snowling, M. (2005). The cognitive foundations of reading and arithmetic skills in 7- to 10-year-olds. *J. Exp. Child Psychol.* 91, 113–136. doi: 10.1016/j.jecp.2005.01.003
- Gallistel, C. R., and Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition* 44, 43–74. doi: 10.1016/0010-0277(92)90050-R
- Geary, D. C., Bow-Thomas, C. C., Liu, F., and Siegler, R. S. (1996). Development of arithmetical competencies in Chinese and American children: influence of age, language, and schooling. *Child Dev.* 67, 2022–2044. doi: 10.2307/1131607
- Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature* 447, 589–591. doi: 10.1038/nature05850
- Gunderson, E. A., Ramirez, G., Beilock, S. L., and Levine, S. C. (2012). The relation between spatial skill and early number knowledge: the role of the linear number line. *Dev. Psychol.* 48, 1229–1241. doi: 10.1037/a0027433
- Halberda, J., Mazocco, M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* 455, 665–668. doi: 10.1038/nature07246
- Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001
- Hubbard, E. M., Piazza, M., Pinel, P., and Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nat. Rev. Neurosci.* 6, 435–448. doi: 10.1038/nrn1684
- Kelly, M. K., Miller, K. F., Fang, G., and Feng, G. (1999). When days are numbered: calendar structure and the development of calendar processing in English and Chinese. *J. Exp. Child Psychol.* 73, 289–314. doi: 10.1006/jecp.1999.2503
- Kolkman, M. E., Kroesbergen, E. H., and Leseman, P. P. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learn. Instr.* 25, 95–103. doi: 10.1016/j.learninstruc.2012.12.001
- Knops, A., Viarouge, A., and Dehaene, S. (2009). Dynamic representations underlying symbolic and nonsymbolic calculation: evidence from the operational momentum effect. *Atten. Percept. Psychophys.* 71, 803–821. doi: 10.3758/APP.71.4.803
- Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., et al. (2011). Mental number line training in children with developmental dyscalculia. *Neuroimage* 57, 782–795. doi: 10.1016/j.neuroimage.2011.01.070
- Landerl, K., Bevan, A., and Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8–9-year-old students. *Cognition* 93, 99–125. doi: 10.1016/j.cognition.2003.11.004

- Laski, E. V., and Siegler, R. S. (2007). Is 27 a big number? Correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Dev.* 78, 1723–1743. doi: 10.1111/j.1467-8624.2007.01087.x
- Laski, E. V., and Yu, Q. (2014). Number line estimation and mental addition: examining the potential roles of language and education. *J. Exp. Child Psychol.* 117, 29–44. doi: 10.1016/j.jecp.2013.08.007
- LeFevre, J.-A., Sadesky, G. S., and Bisanz, J. (1996). Selection of procedures in mental addition: reassessing the problem size effect in adults. *J. Exp. Psychol. Learn. Mem. Cogn.* 22, 216–230. doi: 10.1037/0278-7393.22.1.216
- Linsen, S., Verschaffel, L., Reynvoet, B., and De Smedt, B. (2014). The association between children's numerical magnitude processing and mental multi-digit subtraction. *Acta Psychol.* 145, 75–83. doi: 10.1016/j.actpsy.2013.10.008
- Maertens, B., De Smedt, B., Sasanguie, D., Elen, J., and Reynvoet, B. (2016). Enhancing arithmetic in pre-schoolers with comparison or number line estimation training: does it matter? *Learn. Instr.* 46, 1–11. doi: 10.1016/j.learninstruc.2016.08.004
- McLean, J. F., and Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *J. Exp. Child Psychol.* 74, 240–260. doi: 10.1006/jecp.1999.2516
- McNeil, N. M., Fuhs, M. W., Keultjes, M. C., and Gibson, M. H. (2011). Influences of problem format and SES on preschoolers' understanding of approximate addition. *Cogn. Dev.* 26, 57–71. doi: 10.1016/j.cogdev.2010.08.010
- Menon, V., Rivera, S., White, C., Glover, G., and Reiss, A. (2000). Dissociating prefrontal and parietal cortex activation during arithmetic processing. *Neuroimage* 12, 357–365. doi: 10.1006/nimg.2000.0613
- Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: the role of number-naming systems. *Psychol. Sci.* 6, 56–60. doi: 10.1111/j.1467-9280.1995.tb00305.x
- Moeller, K., Fischer, U., Nuerk, H.-C., and Cress, U. (2015). Computers in mathematics education—Training the mental number line. *Comput. Human Behav.* 48, 597–607. doi: 10.1016/j.chb.2015.01.048
- Mussolin, C., Mejias, S., and Noël, M.-P. (2010). Symbolic and nonsymbolic number comparison in children with and without dyscalculia. *Cognition* 115, 10–25. doi: 10.1016/j.cognition.2009.10.006
- Ng, S. S., and Rao, N. (2010). Chinese number words, culture, and mathematics learning. *Rev. Educ. Res.* 80, 180–206. doi: 10.3102/0034654310364764
- Nys, J., Content, A., and Leybaert, J. (2013). Impact of language abilities on exact and approximate number skills development: evidence from children with specific language impairment. *J. Speech Lang. Hear. Res.* 56, 956–970. doi: 10.1044/1092-4388(2012)10-0229
- Pinel, P., Dehaene, S., Riviere, D., and LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *Neuroimage* 14, 1013–1026. doi: 10.1006/nimg.2001.0913
- Reys, R. E., and Yang, D.-C. (1998). Relationship between computational performance and number sense among sixth- and eighth-grade students in Taiwan. *J. Res. Math. Educ.* 29, 225–237. doi: 10.2307/749900
- Rodic, M., Zhou, X., Tikhomirova, T., Wei, W., Malykh, S., Ismatulina, V., et al. (2015). Cross-cultural investigation into cognitive underpinnings of individual differences in early arithmetic. *Dev. Sci.* 18, 165–174. doi: 10.1111/desc.12204
- Rousselle, L., and Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs non-symbolic number magnitude processing. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005
- Rubinsten, O., Dana, S., Lavro, D., and Berger, A. (2013). Processing ordinality and quantity: ERP evidence of separate mechanisms. *Brain Cogn.* 82, 201–212. doi: 10.1016/j.bandc.2013.04.008
- Sasanguie, D., and Reynvoet, B. (2013). Number comparison and number line estimation rely on different mechanisms. *Psychol. Belg.* 53, 17–35. doi: 10.5334/pb-53-4-17
- Sasanguie, D., Van den Bussche, E., and Reynvoet, B. (2012). Predictors for mathematics achievement? Evidence from a longitudinal study. *Mind Brain Educ.* 6, 119–128. doi: 10.1111/j.1751-228X.2012.01147.x
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20, e12372. doi: 10.1111/desc.12372
- Siegler, R. S., and Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychol. Sci.* 19, 759–763. doi: 10.1111/j.1467-9280.2008.02153.x
- Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–250. doi: 10.1111/1467-9280.02438
- Siegler, R. S., and Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Dev. Sci.* 11, 655–661. doi: 10.1111/j.1467-7687.2008.00714.x
- Spelke, E. S., and Tsivkin, S. (2001). Language and number: a bilingual training study. *Cognition* 78, 45–88. doi: 10.1016/S0010-0277(00)00108-6
- Sweller, J., Van Merriënboer, J. J., and Paas, F. G. (1998). Cognitive architecture and instructional design. *Educ. Psychol. Rev.* 10, 251–296. doi: 10.1023/A:1022193728205
- Tang, J., Ward, J., and Butterworth, B. (2008). Number forms in the brain. *J. Cogn. Neurosci.* 20, 1547–1556. doi: 10.1162/jocn.2008.20120
- Tosto, M. G., Petrill, S. A., Malykh, S., Malki, K., Haworth, C., Mazzocco, M. M., et al. (2017). Number sense and mathematics: which, when and how? *Dev. Psychol.* 53, 1924–1939. doi: 10.1037/dev0000331
- Turconi, E., Jemel, B., Rossion, B., and Seron, X. (2004). Electrophysiological evidence for differential processing of numerical quantity and order in humans. *Cogn. Brain Res.* 21, 22–38. doi: 10.1016/j.cogbrainres.2004.05.003
- Turconi, E., and Seron, X. (2002). Dissociation between order and quantity meanings in a patient with Gerstmann syndrome. *Cortex* 38, 911–914. doi: 10.1016/S0010-9452(08)70069-8
- Van Merriënboer, J. J., and Sweller, J. (2005). Cognitive load theory and complex learning: recent developments and future directions. *Educ. Psychol. Rev.* 17, 147–177. doi: 10.1007/s10648-005-3951-0
- Vanbinst, K., Ghesquière, P., and De Smedt, B. (2015). Does numerical processing uniquely predict first graders' future development of single-digit arithmetic? *Learn. Individ. Differ.* 37, 153–160. doi: 10.1016/j.lindif.2014.12.004
- Von Aster, M. G., and Shalev, R. S. (2007). Number development and developmental dyscalculia. *Dev. Med. Child Neurol.* 49, 868–873. doi: 10.1111/j.1469-8749.2007.00868.x
- Wilson, A. J., Dehaene, S., Dubois, O., and Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind Brain Educ.* 3, 224–234. doi: 10.1111/j.1751-228X.2009.01075.x
- Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., and Dehaene, S. (2006). An open trial assessment of "The Number Race", an adaptive computer game for remediation of dyscalculia. *Behav. Brain Funct.* 2:20.
- Xenidou-Dervou, I., De Smedt, B., van der Schoot, M., and van Lieshout, E. C. (2013). Individual differences in kindergarten math achievement: the integrative roles of approximation skills and working memory. *Learn. Individ. Differ.* 28, 119–129. doi: 10.1016/j.lindif.2013.09.012
- Xenidou-Dervou, I., Gilmore, C., van der Schoot, M., and van Lieshout, E. C. (2015). The developmental onset of symbolic approximation: beyond nonsymbolic representations, the language of numbers matters. *Front. Psychol.* 6:487. doi: 10.3389/fpsyg.2015.00487
- Zacks, J. M. (2008). Neuroimaging studies of mental rotation: a meta-analysis and review. *J. Cogn. Neurosci.* 20, 1–19. doi: 10.1162/jocn.2008.20013
- Zago, L., Pesenti, M., Mellet, E., Crivello, F., Mazoyer, B., and Tzourio-Mazoyer, N. (2001). Neural correlates of simple and complex mental calculation. *Neuroimage* 13, 314–327. doi: 10.1006/nimg.2000.0697
- Zago, L., and Tzourio-Mazoyer, N. (2002). Distinguishing visuospatial working memory and complex mental calculation areas within the parietal lobes. *Neurosci. Lett.* 331, 45–49. doi: 10.1016/S0304-3940(02)00833-9

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Wei, Deng, Chen, He, Qin and Kovas. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



The Role of Approximate Number System in Different Mathematics Skills Across Grades

Dan Cai¹, Linni Zhang¹, Yan Li¹, Wei Wei^{1*} and George K. Georgiou²

¹ College of Education, Shanghai Normal University, Shanghai, China, ² Department of Educational Psychology, University of Alberta, Edmonton, AB, Canada

OPEN ACCESS

Edited by:

Jingguang Li,
Dali University, China

Reviewed by:

Ruomeng Zhao,
MacPractice, Inc., United States
Song Wang,
Sichuan University, China
Kathy Ellen Green,
University of Denver, United States

*Correspondence:

Wei Wei
wwwei@shnu.edu.cn

Specialty section:

This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

Received: 08 April 2018

Accepted: 27 August 2018

Published: 18 September 2018

Citation:

Cai D, Zhang L, Li Y, Wei W and
Georgiou GK (2018) The Role
of Approximate Number System
in Different Mathematics Skills Across
Grades. *Front. Psychol.* 9:1733.
doi: 10.3389/fpsyg.2018.01733

Although approximate number system (ANS) has been found to predict mathematics ability, it remains unclear if both aspects of ANS (symbolic and non-symbolic estimation) contribute equally well to mathematics performance and if their contribution varies as a function of the mathematics outcome and grade level. Thus, in this study, we examined the effects of both aspects of ANS on different mathematics skills across three grade levels. Three hundred eleven children (100 children from kindergarten, 107 children from Grade 2, and 104 children from Grade 4) from two kindergartens and three elementary schools in Shanghai, China, were assessed on measures of ANS (dot estimation and number line estimation), general cognitive ability (nonverbal intelligence, inhibition, and working memory), and mathematics abilities (numerical operations and mathematical problem solving in all grades, early mathematical skills in kindergarten, and calculation fluency in Grades 2 and 4). Results of hierarchical regression analyses showed that, in kindergarten, non-symbolic estimation predicted all mathematics skills even after controlling for age, gender, and general cognitive ability. In Grades 2 and 4, symbolic estimation accounted for unique variance in mathematical problem solving, but not in calculation fluency. Symbolic estimation also predicted numerical operations in Grade 4. Taken together, these findings suggest that in the early phases of mathematics development different aspects of ANS contribute to different mathematics skills.

Keywords: approximate number system, non-symbolic estimation, symbolic estimation, mathematics skills, Chinese

INTRODUCTION

The approximate number system (ANS) is a mental system responsible for representing and processing numerical magnitude information (De Smedt et al., 2013; Libertus, 2015). It has been argued that ANS helps children form imprecise numerical estimations that are later on activated and used in magnitude comparisons (Siegler and Lortie-Forgues, 2014) and in mathematics learning (see Clements and Sarama, 2007; Feigenson et al., 2013; Libertus, 2015; Mussolin et al., 2016, for reviews). However, far less is known about the conditions under which the two most known ANS aspects (symbolic and non-symbolic estimation) predict mathematics skills. Therefore, this study aimed to examine how the two ANS aspects (symbolic and non-symbolic estimation) contribute to different mathematics skills (early mathematics skills, numerical operations, mathematical problem solving, and calculation fluency) in different grade levels (kindergarten, Grade 2, and Grade 4).

Approximate number system consists of two aspects: non-symbolic estimation and symbolic estimation. Non-symbolic estimation refers to the processing of quantities and numerosities

without using numerals (Smets et al., 2015). It emerges as early as at the age of 6 months, when infants discriminate between large ratios of two arrays (e.g., 6:12; Libertus and Brannon, 2010), and continues to develop until adulthood, when individuals use this knowledge to discriminate between smaller ratios (e.g., 0.9:1; Price et al., 2012). In turn, symbolic estimation refers to mapping the numerals on a quantitative dimension, such as approximating the number of dots in a picture and the location of a number on a number line (Booth and Siegler, 2006). It is hypothesized that the numerals are mentally represented along a mental number line (Siegler and Lortie-Forgues, 2014) and the representations of numerals become more accurate from a logarithmic manner to a linear manner as children get older (Siegler and Booth, 2004; Friso-van den Bos et al., 2015). Meta-analyses have reported significant correlations between the two ANS aspects and mathematics (Chen and Li, 2014; Fazio et al., 2014; Schneider et al., 2018a). For example, Chen and Li (2014) estimated the average correlation between non-symbolic estimation and mathematics to be 0.24, and Schneider et al. (2018a) reported an average correlation between symbolic estimation and mathematics of 0.44.

The meta-analyses, however, have also detected great heterogeneity among the correlations. A possible explanation for this heterogeneity may be that the two ANS aspects exert a different effect on mathematics skills in different grades. To delineate this, a study should examine the role of both ANS aspects in mathematics across different grade levels (what we did in our study). Besides, it is also possible that the effects of grade level interact with the type of mathematics skill assessed in different studies. Mathematics skills include a wide range of skills such as early mathematics skills (e.g., counting and number knowledge), numerical operations (i.e., the ability to use algorithms to solve written arithmetic), calculation fluency (the ability to retrieve arithmetic facts from memory quickly), and mathematical problem solving (the ability to apply mathematical concepts and arithmetic to solve contextual problems). Some researchers (Libertus et al., 2013; Wang et al., 2016) have argued that non-symbolic estimation may help children learn number-related knowledge such as number concepts, number intrarelations, and thus be more important for early mathematics abilities. In later years, symbolic estimation may help children understand symbolic arithmetic and facilitate recall of answers to arithmetic problems (Siegler and Braithwaite, 2017), and thus be more important in mathematics in later grades. Recently, Tosto et al. (2017) also argued that once arithmetic skills become automatized, neither non-symbolic nor symbolic estimation should play an important role. This should particularly affect calculation fluency since children (particularly Chinese)¹ become efficient in executing simple calculations as early as in Grade 1 (e.g., Deng et al., 2015; Cui et al., 2017).

Only a few studies have also contrasted the effects of both symbolic and non-symbolic estimation in the same study (e.g., Sasanguie et al., 2012, 2013; Jordan et al., 2013; Lyons et al., 2014;

Cirino et al., 2016; Tosto et al., 2017). Most of these studies have shown that number line estimation uniquely explains mathematics skills after controlling for non-symbolic estimation (e.g., Sasanguie et al., 2012; Jordan et al., 2013; Lyons et al., 2014; Cirino et al., 2016; Tosto et al., 2017), but none of these studies have examined how the two ANS skills explain early mathematics skills. Although non-symbolic estimation appears to be less important in learning mathematics in school years, as reviewed before, it may uniquely explain mathematics skills in early years.

Interestingly, most of the previous studies examining the role of ANS in mathematics did not control for the effects of key cognitive predictors of mathematics such as nonverbal intelligence or executive functioning. Executive functioning, the cognitive skills engaged in goal-directed activities, includes inhibition and working memory (e.g., Miyake et al., 2000; Lehto et al., 2003), both of which are significant correlates of mathematics skills (e.g., Swanson, 2006; Bull et al., 2008; Lan et al., 2011; Cragg et al., 2017; see Bull and Lee, 2014, for a review). Executive functioning may also contribute to non-symbolic and symbolic estimation (e.g., Xenidou-Dervou et al., 2013; Wong et al., 2016; Peng et al., 2017; Zhu et al., 2017; Purpura and Simms, 2018). Inhibition may be required in suppressing non-numerical stimulus features and focus attention on the magnitude (Starr et al., 2017), and working memory may be needed in holding symbolic or non-symbolic information in rapid comparison of two arrays of objects (Xenidou-Dervou et al., 2013) and in holding the bounds or referent points and their corresponding values in number line tasks (Schneider et al., 2018b). Therefore, the association between ANS acuity and mathematics may be accounted for by executive functioning. Price and Wilkey (2017), for example, found that inhibition and working memory partly mediated the relationship between ANS acuity (both non-symbolic and symbolic estimation) and mathematics skills.

Notice also that most previous studies on ANS were conducted in Western countries and far less is known about the role of ANS acuity in learning mathematics in East Asian countries (e.g., China). The place-value system in Chinese is relatively transparent (e.g., “十一(ten-one)” for eleven), which may facilitate Chinese children learning symbolic numbers (Miller et al., 2005). The easier mastery of symbolic numbers in Chinese may result in non-symbolic estimation being less important in learning mathematics. There are reasons to believe that non-symbolic and symbolic estimation may play a different role in China than in Western countries. To date, only a handful of studies have examined the effects of symbolic or non-symbolic estimation on mathematics in Chinese children (see Lonnemann et al., 2011; He et al., 2016; Wang et al., 2016; Wong et al., 2016; Zhang et al., 2016; Cui et al., 2017; Peng et al., 2017; Zhu et al., 2017), and none of these studies have examined how symbolic and non-symbolic estimation predict different mathematics skills in both early and later elementary school years.

Therefore, the present study aimed to examine the effects of both ANS aspects (symbolic and non-symbolic estimation) on different mathematics skills (early mathematics skills, numerical operations, mathematical problem solving, and calculation fluency) in different grade levels in China. Based on the findings

¹This is because Chinese children attend Kindergarten at the age of 3 and stay in kindergarten for 3 years before they go to Grade 1. In Kindergarten, they learn to perform simple calculations.

of previous studies (Jordan et al., 2013; Lyons et al., 2014; Wong et al., 2016; Tosto et al., 2017; Zhu et al., 2017), we hypothesized that:

- 1) The effects of symbolic and non-symbolic estimation will vary as a function of grade level. Non-symbolic estimation will uniquely predict mathematics skills only in kindergarten, while symbolic estimation will uniquely predict mathematics skills at all grade levels.
- 2) Non-symbolic and symbolic estimation will predict different mathematics skills. Non-symbolic estimation will predict early mathematics skills, that is counting, symbolic number knowledge and arithmetic, and symbolic estimation will predict all mathematics skills except from calculation fluency.

MATERIALS AND METHODS

Participants

The participants were 100 children from kindergarten (53 girls and 47 boys; mean age = 66.53 months, $SD = 3.31$), 107 children from Grade 2 (60 girls and 47 boys; mean age = 92.16 months, $SD = 3.96$), and 104 children from Grade 4 (59 girls and 44 boys; mean age = 115.75 months, $SD = 3.62$). The children were recruited on a voluntary basis from two kindergartens and three elementary schools in Shanghai, China. The schools that participated in our study serve primarily middle-class families and the demographics are representative of the general population in Shanghai (The National Bureau of Statistics in Shanghai, 2017). All children were native Mandarin speakers and none was diagnosed with any intellectual, sensory, or behavioral disorders. Parental consent and ethics approval from the Shanghai Normal University were obtained prior to testing.

Materials

General Cognitive Abilities

Nonverbal intelligence

Nonverbal Matrices from Cognitive Assessment System-Version 2 (CAS-2; Naglieri et al., 2014), was used to assess nonverbal intelligence. Children were presented with a variety of geometric designs that were missing one part and were asked to select the missing part among six options. The task was discontinued after four consecutive errors. The score was the total number correct (max = 44). Criterion validity has been reported to range from 0.57 to 0.65 (Naglieri et al., 2014). The Cronbach's alpha reliability coefficient in the current study was 0.85 in Kindergarten, 0.91 in Grade 2, and 0.90 in Grade 4.

Executive functioning

Inhibition. Expressive Attention, adopted from CAS-2 (Naglieri et al., 2014) was used to assess children's inhibition. Two versions (5–7 years and 8–18 years) were used to avoid ceiling/floor effects. The version used for children in Grades 2 and 4 is similar to the color-word Stroop test (Stroop, 1935) and includes three pages. In the first page, children were asked to say aloud the names of color squares (e.g., blue, yellow, red, and green) and, in the

second, children were asked to name the color characters (e.g., “黄,” yellow). In the third page, children were presented with 40 color characters each printed in a color different from the color character [e.g., “黄 (yellow)” printed in blue ink]. They were asked to read aloud the color of the ink in which the characters were printed as quickly as possible. An 8-item practice trial was used to make sure children understood the instructions prior to testing. A ratio score was calculated by dividing the number of correct responses by the time to finish naming all 40 items. Criterion validity has been reported to range from 0.69 to 0.73 (Naglieri et al., 2014). The Cronbach's alpha reliability coefficient in the current study was 0.86 in both Grades 2 and 4.

The version for 5–7 year old students was used in kindergarten and it also included three pages. In each page, children were shown animal drawings that included small animals (butterfly, mouse, bird, and frog) and big animals (elephant, whale, horse, and bear), and were asked to say aloud whether each animal was small or big as fast as they could. In the first page, animal drawings were printed in a uniform size, and in the second page, big animals were printed in big size and small animals in a small size. In the third page, big animals were printed in a small size and small animals in a big size, and children were asked to name the animal drawing based on their actual size and not based on the size they were printed. The score was the number of correct responses in the third page divided by the time to finish naming the items. Criterion validity has been reported to range from 0.51 to 0.67 (Naglieri et al., 2014). The Cronbach's alpha reliability coefficient in the current study was 0.81.

Working memory. Digit Span Forward from CAS-2 (Naglieri et al., 2014) was used to assess children's working memory. The test consists of 2–9 span with four trials in each span. The numbers were orally presented at the speed of one number per second and then children were asked to repeat these numbers in the same order. The test was discontinued when three errors were made in each span. The score was the final span that the children had reached. Criterion validity has been reported to range from 0.40 to 0.64 (Naglieri et al., 2014). The Cronbach's alpha reliability coefficient in the current study was 0.88, 0.89, and 0.88 in Kindergarten, Grade 2, and Grade 4, respectively.

Mathematics Skills

Early mathematics skills

Test of early mathematics ability (TEMA-3; Ginsburg and Baroody, 2003) was used to measure kindergarteners' early mathematics skills. TEMA-3 includes 72 items on counting, symbolic number knowledge, and arithmetic. The test was discontinued after four consecutive errors and the children's score was the total number correct. TEMA-3 has been found to correlate significantly with other math tests such as Mathematics subtest of the Young Children's Achievement Test and Key Math Revised (r 's range from 0.54 to 0.91; Ginsburg and Baroody, 2003). The Cronbach's alpha reliability coefficient in the current study was 0.88.

Numerical operations

Numerical operations, adopted from WIAT-III (Wechsler Individual Achievement Test-Third Edition; Wechsler, 2009),

was used to assess children's numerical operations skills under untimed conditions. The items were arranged in increasing difficulty and children were asked to solve these items one by one. The test was discontinued after four consecutive errors and a participant's score was the total number correct. Numerical operations has been found to correlate significantly with other math measures such as numerical operations in WIAT-II and Math Reasoning (r 's range from 0.71 to 0.81; Wechsler, 2009). The Cronbach's alpha reliability coefficient in the current study was 0.80 and 0.89 in Grades 2 and 4, respectively.

Mathematical problem solving

Math problem solving, adopted from WIAT-III (Wechsler Individual Achievement Test-Third Edition; Wechsler, 2009), was used to assess mathematical problem solving. The items in the task were arranged in terms of increasing difficulty (max = 72). Children were asked to solve these items one by one, under untimed conditions. The test was discontinued after four consecutive errors and a participant's score was the total number correct. Math problem solving has been found to correlate significantly with other math measures such as numerical operations and math reasoning (r 's range from 0.75 to 0.84; Wechsler, 2009). The Cronbach's alpha reliability coefficient in the current study was 0.88, 0.90, and 0.90 in Kindergarten, Grade 2, and Grade 4, respectively.

Calculation fluency

Math fluency from WIAT-III (Wechsler Individual Achievement Test-Third Edition; Wechsler, 2009) was used to assess children's calculation fluency. This task includes three subtests: addition fluency (e.g., $5 + 1 = 6$), subtraction fluency (e.g., $4 - 2 = 2$), and multiplication fluency (e.g., $2 \times 3 = 6$). Children were asked to write down the answers to 48 items in each subtest as soon as they could in 1-min time limit. A participant's score was the sum of three subtests' scores. Math fluency has been found to correlate significantly with other math measures such as numerical operations and math reasoning (r 's range from 0.55 to 0.64; Wechsler, 2009). Zhu et al. (2017) reported internal consistency reliability for math fluency to be 0.88 and 0.93 for Grades 2 and 4, respectively.

Approximate Number System

Non-symbolic estimation

Dot estimation, adapted from Halberda and Feigenson (2008), was used to assess non-symbolic estimation task on a computer. At the time of testing, two pictures would appear on the screen. There were some random points (10–30 points) on each picture. The number of points on the two pictures was different. In Grades 2 and 4, children were asked to judge which picture had more points within a 2 s time limit. In kindergarten, children were given 3 s to make a decision². The task included 6 practice items and 24 test items. A participant's score was the percentage of accurate responses across the 24 items. The task has been used in several previous studies in Chinese showing good psychometric

properties (e.g., Cui et al., 2017; Zhu et al., 2017; Cheng et al., 2018). The Cronbach's alpha reliability coefficient in the current study was 0.69, 0.77, and 0.72, for Kindergarten, Grade 2, and Grade 4, respectively.

Symbolic estimation

Number line estimation was adopted from Opfer and Siegler (2007) and was used to measure children's symbolic estimation. The version for Grade 2 and Grade 4 was carried out on an 8-inch tablet. There was a line displayed on the pad 0 was marked on the left of the line, and 100 was marked on the right. At the time of testing, a number would appear on the screen, and children were asked to estimate which position this number was in 0–100 and mark the position on the line. The items included 26 numbers: 3, 4, 6, 8, 12, 17, 20, 21, 23, 25, 29, 33, 39, 43, 48, 50, 52, 57, 61, 64, 72, 79, 81, 84, 90, and 96. The items were presented in random order. In kindergarten, the number line task was given as a paper and pencil task. The actual length of the line was 24 cm and it was used to represent the distance from 0 to 10. The items included nine numbers: 1, 2, 3, 4, 5, 6, 7, 8, and 9. The formula to calculate the final score was: $\frac{\text{Estimation} - \text{Estimation Quantity}}{\text{Scale of Estimation}}$. The task has been used in previous studies in Chinese showing good psychometric properties (e.g., Siegler and Mu, 2008; Laski and Yu, 2014; Zhu et al., 2017). The Cronbach's alpha reliability coefficient in our sample was 0.72, 0.80, and 0.69, for Kindergarten, Grade 2, and Grade 4, respectively.

Procedures

Children were individually tested by trained graduate students in a quiet room in their school. The testing was completed in two sessions of 30–40 min each. Session A included the mathematics tests [math problem solving, numerical operations, math fluency (only in Grades 2 and 4), TEMA-3 (only in Kindergarten)]. Session B included the cognitive tests (nonverbal matrices, expressive attention, and digit span forward) and the ANS tasks (dot estimation and number line estimation). Half of the children in each grade level did first Session A and then Session B. The other half did the sessions in the reverse order.

RESULTS

Preliminary Data Analyses

Table 1 shows the descriptive statistics (mean, standard deviation, range, and kurtosis and skewness) for all the measures in our study. The distributions of numerical operations and dot estimation were positively skewed and thus log transformation was applied. After the log transformation, their distributions were normalized and the transformed scores were used in further analyses.

Correlations Between the Measures

The correlation coefficients among all variables in kindergarten, Grade 2, and Grade 4 are presented in Tables 2, 3. In kindergarten, both the number line estimation and dot estimation correlated significantly with all mathematics skills (r 's ranged from -0.43 to -0.55). In Grade 2, number line estimation

²This time limit was decided based on a pilot study we conducted as well as based on the time limit used in previous studies with children of the same age as ours (e.g., Fazio et al., 2014; Libertus et al., 2016).

TABLE 1 | Descriptive Statistics for all Measures Used in our Study.

	Kindergarten				Grade 2				Grade 4			
	M (SD)	Range	Skew.	Kurt.	M (SD)	Range	Skew.	Kurt.	M (SD)	Range	Skew.	Kurt.
NO	14.66 (4.98)	0–26	−0.78	0.96	20.83 (2.84)	16–30	1.14	1.18	33.41 (3.21)	24–52	2.11	11.48
NO (lg)	1.14 (0.17)	0.67–1.41	−1.46	1.60	1.32 (0.06)	1.2–1.5	0.78	0.55	1.52 (0.03)	1.45–1.63	0.92	1.79
MPS	30.52 (5.65)	13–41	−0.53	0.19	42.35 (4.67)	31–58	0.08	0.71	50.61 (3.52)	42–65	1.16	3.55
TEMA	42.79 (13.02)	16–69	−0.03	−0.69	\	\	\	\	\	\	\	\
MF	\	\	\	\	67.21 (14.07)	39–98	0.15	−0.76	106.79 (16.96)	71–136	−0.01	−0.78
DE	0.78 (0.13)	0.38–0.96	−1.00	0.69	0.73 (0.10)	0.52–0.96	0.10	−0.23	0.77 (0.11)	0.38–0.96	−1.23	2.58
NLE	0.19 (0.09)	0.04–0.59	0.96	3.15	0.07 (0.03)	0.02–0.21	1.58	3.70	0.06 (0.03)	0.02–0.26	3.60	19.49
NLE(lg)	−0.77 (0.23)	−1.44–(−0.23)	−0.80	0.51	−1.20 (0.18)	−1.69–(−0.69)	0.20	0.19	−0.26 (0.17)	−1.75–(−0.59)	0.62	2.99
Intelligence	18.44 (6.15)	1–32	−0.66	0.76	25.12 (5.00)	14–37	0.19	−0.45	26.91 (5.89)	16–39	0.36	−0.73
Inhibition	0.64 (0.17)	0.03–1.09	−0.16	1.72	1.25 (0.47)	0.37–2.50	0.02	−0.49	0.74 (0.19)	0.08–1.31	−0.10	1.46
WM	6.75 (1.46)	4–9	−0.05	−1.07	8.00 (1.03)	5–9	−0.81	−0.23	8.23 (0.90)	6–9	−0.97	0.10

NO, numerical operations; MPS, math problem solving; TEMA, test of early mathematics ability; MF, math fluency; DE, dot estimation; NLE, number line estimation; Intelligence, nonverbal intelligence; WM, working memory.

correlated significantly with math problem solving ($r = -0.52$) and math fluency ($r = -0.21$). In Grade 4, number line estimation correlated significantly with math problem solving ($r = -0.28$) and numerical operations ($r = -0.27$). Dot estimation did not correlate significantly with any math task in Grades 2 and 4.

Results of Regression Analyses

Hierarchical regression analyses were subsequently conducted within each grade level to examine the unique contribution of the two ANS aspects to mathematics outcomes [math problem solving, numerical operations, math fluency (assessed only in Grades 2 and 4), and TEMA (assessed only in kindergarten)]. In each model, age and gender were entered in the regression equation at step 1 as control variables. The general cognitive abilities (nonverbal intelligence, inhibition, and working memory) were entered in the regression equation at step 2, and number line estimation and dot estimation were entered at step 3 of the regression equation as a block.

Tables 4–6 show the standardized beta coefficients, R^2 changes, and significance levels of the regression models in each grade level. In kindergarten, the two ANS aspects accounted for unique variance in math problem solving [5%, but only dot estimation had a significant effect ($\beta = -0.190$, $p < 0.01$)], numerical operations [4%, but only dot estimation had a significant effect ($\beta = -0.192$, $p < 0.05$)], and TEMA-3 [17%, both number line estimation ($\beta = -0.358$, $p < 0.001$) and dot estimation ($\beta = -0.246$, $p < 0.01$) had a significant effect], after controlling for age, gender, nonverbal intelligence, inhibition, and working memory. In Grade 2, ANS accounted for unique variance in math problem solving [14%, but only the effects of number line estimation were significant ($\beta = -0.444$, $p < 0.001$)], but not in numerical operation and math fluency. In Grade 4, ANS accounted for unique variance in math problem solving [5%, but only the effects of number line estimation were significant ($\beta = -0.184$, $p < 0.05$)], but not in math fluency. The predictive effect of number line estimation on numerical operations was also significant ($\beta = -0.203$, $p < 0.05$).

DISCUSSION

The aim of this study was to examine how two ANS aspects (symbolic and non-symbolic estimation) predict different mathematics skills in different grade levels in China. Overall, our findings showed that the relationship between ANS acuity and mathematics skills depends on the type of ANS aspect, the type of mathematics outcome assessed, and the grade level. Among kindergarteners, non-symbolic estimation uniquely predicted early mathematics skills, numerical operations, and mathematical problem solving. Symbolic estimation explained unique variance only in early mathematics skills. Symbolic estimation also predicted mathematical problem solving among the second- and fourth-graders, and numerical operations among the fourth-graders.

In line with our expectation, non-symbolic estimation made unique contributions to mathematics skills only in kindergarten. This replicates the findings of earlier studies, which found that non-symbolic estimation played a unique role in early mathematics skills (e.g., Clements and Sarama, 2007; Inglis et al., 2011; Desoete et al., 2012; Xenidou-Dervou et al., 2016; Starr et al., 2017). As Xenidou-Dervou et al. (2016) have noted, the start of formal mathematics education may cause symbolic estimation to become a prominent predictor of mathematics skills. It should be noted that non-symbolic estimation in kindergarten made a substantial contribution to early mathematics skills other than numerical operations and mathematical problem solving, which replicates the results of a recent meta-analysis (Schneider et al., 2017). Schneider et al. (2017) found that the correlation between non-symbolic estimation and early mathematics skills was higher than that between non-symbolic estimation and formal mathematics skills such as arithmetic. Previous studies have also shown that non-symbolic estimation correlates highly with early numerical skills such as counting and non-symbolic arithmetic (Gillmore et al., 2007; Libertus et al., 2013; van Marle et al., 2014).

Symbolic estimation made unique contributions to mathematical problem solving in Grades 2 and 4, and to

TABLE 2 | Correlations between the variables in kindergarten.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Math problem solving										
2. Numerical operation	0.73**									
3. TEMA	0.84**	0.68**								
4. Number Line Estimation	−0.37**	−0.28**	−0.55**							
5. Dot Estimation	0.40**	0.37**	0.43**	−0.13						
6. Nonverbal Intelligence	0.68**	0.63**	0.62**	−0.33**	0.28**					
7. Inhibition	0.33**	0.37**	0.34**	−0.07	0.25*	0.25*				
8. Working Memory	0.37**	0.22*	0.36**	−0.19	0.12	0.23*	0.21*			
9. Age	0.21*	0.15	0.13	−0.15	0.13	0.15	0.21*	0.08		
10. Gender	0.12	0.16	0.16	−0.18	−0.09	0.07	0.05	0.07	0.16	

* $p < 0.05$, ** $p < 0.01$.**TABLE 3 |** Correlations between the variables in Grade 2 (below the diagonal) and Grade 4 (above the diagonal).

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.
1. Math problem solving		0.56**	0.28**	−0.28**	−0.09	0.38**	0.21*	0.09	0.06	0.23**
2. Numerical operation	0.19*		0.22*	−0.27**	0.01	0.26**	0.27**	0.15	0.10	0.15
3. Math fluency	0.23*	−0.07		−0.09	−0.00	−0.05	0.24*	0.09	0.06	0.15
4. Number line estimation	−0.52**	−0.14	−0.21*		0.04	−0.21*	−0.08	−0.04	−0.08	−0.09
5. Dot estimation	0.15	0.10	0.05	−0.28**		−0.01	0.16	−0.02	0.16	0.04
6. Nonverbal intelligence	0.26**	0.10	0.01	−0.32**	0.04		0.14	−0.00	−0.02	0.18
7. Inhibition	0.27**	0.12	0.12	−0.26**	0.08	−0.05		−0.07	0.10	0.02
8. Working memory	0.29**	−0.08	0.21*	−0.25*	0.21*	0.28**	0.00		0.04	0.02
9. Age	0.16	0.12	0.12	−0.13	−0.04	0.18	0.25**	−0.03		0.00
10. Gender	0.07	0.16	0.03	−0.13	−0.12	0.07	−0.01	−0.08	0.18	

* $p < 0.05$, ** $p < 0.01$.**TABLE 4 |** Results of hierarchical regression analyses predicting math problem solving in kindergarten, Grade 2, and Grade 4.

Step		Kindergarten		Grade 2		Grade 4	
		β	ΔR^2	β	ΔR^2	β	ΔR^2
1.	Age	0.200*	0.05	0.154	0.03	0.060	0.07*
	Gender	−0.088		−0.030		−0.248*	
2.	Nonverbal intelligence	0.592***	0.49***	0.187	0.17***	0.340***	0.16**
	Inhibition	0.116		0.269**		0.154	
	Working memory	0.205**		0.235*		0.099	
3.	Number line estimation	−0.121	0.05**	−0.444***	0.14***	−0.184*	0.05*
	Dot estimation	0.190**		−0.052		−0.131	
Total R^2			0.59***		0.34***		0.28***

* $p < 0.05$, ** $p < 0.001$, *** $p < 0.001$.

numerical operations in Grade 4. The effect of number line estimation on numerical operations and mathematical problem solving is in line with the findings of previous studies (e.g., Jordan et al., 2013; Tosto et al., 2017; Zhu et al., 2017). It was surprising that symbolic estimation did not uniquely explain numerical operations in Grade 2, although it is in line with Geary (2011), who found that number line estimation in Grade 1 did not concurrently predict numerical operations. It suggests that symbolic estimation may be more important in learning more complex arithmetic such as fractions. Grade 4 students in Chinese are learning fractions (Shanghai Municipal

Education Commission, 2004), and thus are handling fraction problems in the numerical operations task. Previous studies have found that number line estimation is very important in learning fraction knowledge (Jordan et al., 2013; Hansen et al., 2015), since it may provide children with an advantage in learning fraction concepts. Jordan et al. (2013) also argued that fraction knowledge may facilitate the number line estimation since children may use proportion strategies in number line task, such as mentally dividing the line into quarters to get more precise estimation (Siegler and Opfer, 2003).

TABLE 5 | Results of hierarchical regression analyses predicting numerical operations in kindergarten, Grade 2, and Grade 4.

		Kindergarten		Grade 2		Grade 4	
Step		β	ΔR^2	β	ΔR^2	β	ΔR^2
1.	Age	0.123	0.04	0.096	0.04	0.127	0.04
	Gender	-0.140		-0.143		-0.148	
2.	Nonverbal intelligence	0.555***	0.41***	0.127	0.03	0.231*	0.14**
	inhibition	0.212*		0.124		0.240*	
	Working memory	0.045		-0.098		0.157	
3.	Number line estimation	-0.054	0.04*	-0.065	0.01	-0.203*	0.04
	Dot estimation	0.192*		0.093		-0.055	
Total R^2			0.49***		0.08		0.22**

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.**TABLE 6 |** Results of hierarchical regression analyses predicting TEMA and math fluency (MF) in kindergarten, Grade 2, and Grade 4.

		Kindergarten (TEMA-3)		Grade 2 (MF)		Grade 4 (MF)	
Step		β	ΔR^2	β	ΔR^2	β	ΔR^2
1.	Age	0.103	0.04	0.121	0.02	0.021	0.02
	Gender	-0.148		0.004		-0.133	
2.	Nonverbal intelligence	0.525***	0.43***	-0.050	0.06	-0.151	0.09*
	inhibition	0.170*		0.098		0.271**	
	Working memory	0.194*		0.241*		0.104	
3.	Number line estimation	-0.358***	0.17***	-0.148	0.02	-0.067	0.01
	Dot estimation	0.246**		-0.002		-0.033	
Total R^2			0.64***		0.10		0.12*

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In contrast to our expectation, symbolic estimation uniquely explained only TEMA-3, but not numerical operations or mathematical problem solving among kindergartners. This might be due to the fact that early mathematics tasks included items such as number comparison, and number knowledge is closely connected with the performance on number line estimation. Children in kindergarten were learning to map symbolic digits onto pre-existing non-symbolic representations (Barth et al., 2005; Mundy and Gilmore, 2009), and thus the number line estimation correlated with the early mathematics skills. Another reason may be that the early mathematics skills may promote the performance on number line tasks. Previous studies showed that young children typically use counting-based strategies when placing a number on the number line (Petitto, 1990; Schneider et al., 2008), and thus children with better counting skills may estimate more precisely on the number line task.

Symbolic estimation did not uniquely predict calculation fluency among school-age children, which was in line with the findings of previous studies (Sasanguie et al., 2013; Zhu et al., 2017). For example, Sasanguie et al. (2013) found that number line estimation among Grades 1–3 children uniquely predicted their performance on a comprehensive mathematics achievement test 1 year later, but failed to predict their performance on a timed arithmetic test. However, Zhu et al. (2017) found that number line estimation in Grade 2 and not in

Grade 4 uniquely predicted concurrent calculation fluency after controlling for general cognitive abilities. A possible explanation might be that Zhu et al. (2017) did not include non-symbolic estimation in their study. An alternative explanation may be that we used the accuracy of number line estimation, while calculation fluency assessed the speed of arithmetic, which may tap on the speed of activating number representations. Holloway and Ansari (2009) found that the distance effect in a symbolic comparison task (calculated from accuracy scores of elementary children) did not correlate with calculation fluency, while that calculated from the response time scores uniquely explained calculation fluency. As Tosto et al. (2017) have argued, the limited role of symbolic estimation in calculation fluency may indicate that symbolic estimation may be less important for arithmetic once calculation reached an automatic level.

Some limitations of the present study are worth mentioning. First, the cross-sectional design of this study does not allow us to draw conclusions about the causal relationships between the two ANS aspects and mathematics skills. The direction of their relation should be examined further since recent studies also showed that mathematics skills may enhance ANS acuity (e.g., Friso-van den Bos et al., 2015). Second, we did not assess the role of home numeracy environment in our study. Previous studies have found that home numeracy environment is an important predictor of children's mathematics achievement

(e.g., Manolitsis et al., 2013; Deng et al., 2015), and the mathematics activities at home may also promote children's non-symbolic and symbolic estimation (e.g., Muta-Yildiz et al., 2018). Future studies should examine the effects of home numeracy environment on ANS acuity and mathematics skills.

CONCLUSION

Taken together, our results showed that the two ANS aspects have different effects on mathematics skills at different learning periods: non-symbolic estimation was uniquely related to mathematics skills in kindergarten, while symbolic estimation was uniquely related to mathematics skills in elementary school years. These results suggest that different types of ANS acuity should be used to predict mathematic skills in different learning periods and perhaps to identify children at-risk for mathematics difficulties. Moreover, interventions to promote children's mathematics skills should target different ANS aspects for young and school-age children.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of Ethical Guidelines for the Protection of Human Subjects of Research, Academic Ethics Committee at

Shanghai Normal University. The protocol was approved by Academic Ethics Committee at Shanghai Normal University. The parents of all children gave their written consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

DC, GG, WW, and YL designed the study. WW, DC, and LZ collected the data, prepared the data for analysis, and wrote the manuscript. GG, DC, WW, and YL revised the manuscript.

FUNDING

This study was supported by a grant from the National Natural Science Foundation of China (Grant No. 31600906), a grant from the General Project of Shanghai Municipal Education Commission (C16011), and a grant from the China Institute at the University of Alberta.

ACKNOWLEDGMENTS

We would like to thank Zhang Meixia, Luo Qin, Su Hong-Ying, Liang Dandan, and Zha Ling at Shanghai Normal University for their assistance with the data collection.

REFERENCES

- Barth, H., La Mont, K., Lipton, J., and Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14116–14121. doi: 10.1073/pnas.0505512102
- Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42, 189–201. doi: 10.1037/0012-1649.41.6.189
- Bull, R., Espy, K. A., and Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: longitudinal predictors of mathematical achievement at age 7 years. *Dev. Neuropsychol.* 33, 205–228. doi: 10.1080/87565640801982312
- Bull, R., and Lee, K. (2014). Executive functioning and mathematics achievement. *Child Dev. Perspect.* 8, 36–41. doi: 10.1111/cdep.12059
- Chen, Q., and Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychol.* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Cheng, D., Xiao, Q., Chen, Q., Cui, J., and Zhou, X. (2018). Dyslexia and dyscalculia are characterized by common visual perception deficits. *Dev. Neuropsychol.* 43, 497–507. doi: 10.1080/87565641.2018.1481068
- Cirino, P. T., Tolar, T. D., Fuchs, L. S., and Huston-Warren, E. (2016). Cognitive and numerosity predictors of mathematical skills in middle school. *J. Exp. Child Psychol.* 145, 95–119. doi: 10.1016/j.jecp.2015.12.010
- Clements, D. H., and Sarama, J. (2007). "Early childhood mathematics learning," in *Second Handbook of Research on Mathematics Teaching and Learning*, Vol. 1, ed. F. K. Lester, Jr (New York, NY: Information Age Publishing), 461–555.
- Cragg, L., Keeble, S., Richardson, S., Roome, H. E., and Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition* 162, 12–26. doi: 10.1016/j.cognition.2017.01.014
- Cui, J., Georgiou, G. K., Zhang, Y., Li, Y., Shu, H., and Zhou, X. (2017). Examining the relationship between rapid automatized naming and arithmetic fluency in Chinese kindergarten children. *J. Exp. Child Psychol.* 154, 146–163. doi: 10.1016/j.jecp.2016.10.008
- De Smedt, B., Noël, M. P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends Neurosci. Educ.* 2, 48–55. doi: 10.1016/j.tine.2013.06.001
- Deng, C-P., Silinskas, G., Wei, W., and Georgiou, G. K. (2015). Cross-lagged relationships between home learning environment and academic achievement in Chinese. *Early Child. Res. Q.* 33, 12–20. doi: 10.1016/j.jecresq.2015.05.001
- Desoete, A., Ceulemans, A., De Weert, F., and Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *Br. J. Educ. Psychol.* 82, 64–81. doi: 10.1348/2044-8279.002002
- Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53–72. doi: 10.1016/j.jecp.2014.01.013
- Feigenson, L., Libertus, M. E., and Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Dev. Perspect.* 7, 74–79. doi: 10.1111/cdep.12019
- Friso-van den Bos, I., Kroesbergen, E. H., Van Luit, J. E., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., et al. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *J. Exp. Child Psychol.* 134, 12–29. doi: 10.1016/j.jecp.2015.02.002
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Dev. Psychol.* 47, 1539–1552. doi: 10.1037/a0025510
- Gilmore, C. K., McCarthy, S. E., and Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature* 447, 589–591. doi: 10.1038/nature05850
- Ginsburg, H. P., and Baroody, A. J. (2003). *TEMA: Test of Early Mathematics Ability*, 3rd Edn. Austin, TX: PRO-ED.
- Halberda, J., and Feigenson, L. (2008). Developmental change in the acuity of the "Number Sense": the Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Dev. Psychol.* 44, 1457–1465. doi: 10.1037/a0012682

- Hansen, N., Jordan, N. C., Fernandez, E., Siegler, R. S., Fuchs, L., Gersten, R., et al. (2015). General and math-specific predictors of sixth-graders' knowledge of fractions. *Cogn. Dev.* 35, 34–49. doi: 10.1016/j.cogdev.2015.02.001
- He, Y., Zhou, X., Shi, D., Song, H., Zhang, H., and Shi, J. (2016). New evidence on causal relationship between approximate number system (ANS) acuity and arithmetic ability in elementary-school students: a longitudinal cross-lagged analysis. *Front. Psychol.* 7:1052. doi: 10.3389/fpsyg.2016.01052
- Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001
- Inglis, M., Attridge, N., Batchelor, S., and Gilmore, C. (2011). Non-verbal number acuity correlates with symbolic mathematics achievement: but only in children. *Psychon. Bull. Rev.* 18, 1222–1229. doi: 10.3758/s13423-011-0154-1
- Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., and Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *J. Exp. Child Psychol.* 116, 45–58. doi: 10.1016/j.jecp.2013.02.001
- Lan, X., Legare, C. H., Ponitz, C. C., Li, S., and Morrison, F. J. (2011). Investigating the links between the subcomponents of executive function and academic achievement: a cross-cultural analysis of Chinese and American preschoolers. *J. Exp. Child Psychol.* 108, 677–692. doi: 10.1016/j.jecp.2010.11.001
- Laski, E. V., and Yu, Q. (2014). Number line estimation and mental addition: examining the potential roles of language and education. *J. Exp. Child Psychol.* 117, 29–44. doi: 10.1016/j.jecp.2013.08.007
- Lehto, J. E., Juujärvi, P., Kooistra, L., and Pulkkinen, L. (2003). Dimensions of executive functioning: evidence from children. *Br. J. Dev. Psychol.* 21, 59–80. doi: 10.1348/026151003321164627
- Libertus, M. E. (2015). The role of intuitive approximation skills for school math abilities. *Mind Brain Educ.* 9, 112–120. doi: 10.1111/mbe.12072
- Libertus, M. E., and Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Dev. Sci.* 13, 900–906. doi: 10.1111/j.1467-7687.2009.00948.x
- Libertus, M. E., Feigenson, L., and Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learn. Individ. Dif.* 25, 126–133. doi: 10.1016/j.lindif.2013.02.001
- Libertus, M. E., Odic, D., Feigenson, L., and Halberda, J. (2016). The precision of mapping between number words and the approximate number system predicts children's formal math abilities. *J. Exp. Child Psychol.* 150, 207–226. doi: 10.1016/j.jecp.2016.06.003
- Lonnemann, J., Linkersdörfer, J., Hasselhorn, M., and Lindberg, S. (2011). Symbolic and non-symbolic distance effects in children and their connection with arithmetic skills. *J. Neurolinguistics* 24, 583–591. doi: 10.1016/j.jneuroling.2011.02.004
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., and Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Dev. Sci.* 17, 714–726. doi: 10.1111/desc.12152
- Manolitsis, G., Georgiou, G. K., and Tziraki, N. (2013). Examining the effects of home literacy and numeracy environment on early reading and math acquisition. *Early Child. Res. Q.* 28, 692–703. doi: 10.1016/j.ecresq.2013.05.004
- Miller, K. F., Kelly, M., and Zhou, X. (2005). “Learning mathematics in China and the United States: cross-cultural insights into the nature and course of preschool mathematical development,” in *Handbook of Mathematical Cognition*, ed. J. I. D. Campbell (New York, NY: Psychology Press), 163–177.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., and Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: a latent variable analysis. *Cogn. Psychol.* 41, 49–100. doi: 10.1006/cogp.1999.0734
- Mundy, E., and Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *J. Exp. Child Psychol.* 103, 490–502. doi: 10.1016/j.jecp.2009.02.003
- Mussolin, C., Nys, J., Leybaert, J., and Content, A. (2016). How approximate and exact number skills are related to each other across development: a review. *Dev. Rev.* 39, 1–15. doi: 10.1016/j.dr.2014.11.001
- Mutaf-Yildiz, B., Sasanguie, D., De Smedt, B., and Reynvoet, B. (2018). Frequency of home numeracy activities is differentially related to basic number processing and calculation skills in kindergartners. *Front. Psychol.* 9:340. doi: 10.3389/fpsyg.2018.00340
- Naglieri, J. A., Das, J. P., and Goldstein, S. (2014). *CAS2: Cognitive Assessment System*. Austin, TX: Pro-Ed.
- Opfer, J. E., and Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cogn. Psychol.* 55, 169–195. doi: 10.1016/j.cogpsych.2006.09.002
- Peng, P., Yang, X., and Meng, X. (2017). The relation between approximate number system and early arithmetic: the mediation role of numerical knowledge. *J. Exp. Child Psychol.* 157, 111–124. doi: 10.1016/j.jecp.2016.12.011
- Petitto, A. L. (1990). Development of number line and measurement concepts. *Cogn. Instr.* 7, 55–78. doi: 10.1207/s1532690xci0701_3
- Price, G. R., Palmer, D., Battista, C., and Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychol.* 140, 50–57. doi: 10.1016/j.actpsy.2012.02.008
- Price, G. R., and Wilkey, E. D. (2017). Cognitive mechanisms underlying the relation between nonsymbolic and symbolic magnitude processing and their relation to math. *Cogn. Dev.* 44, 139–149. doi: 10.1016/j.cogdev.2017.09.003
- Purpura, D. J., and Simms, V. (2018). Approximate number system development in preschool: what factors predict change? *Cogn. Dev.* 45, 31–39. doi: 10.1016/j.cogdev.2017.11.001
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: what underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418–431. doi: 10.1016/j.jecp.2012.10.012
- Sasanguie, D., Van den Bussche, E., and Reynvoet, B. (2012). Predictors for mathematics achievement? Evidence from a longitudinal study. *Mind Brain Educ.* 6, 119–128. doi: 10.1111/j.1751-228X.2012.01147.x
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372
- Schneider, M., Heine, A., Thaler, V., Torbeyns, J., De Smedt, B., Verschaffel, L., et al. (2008). A validation of eye movements as a measure of elementary school children's developing number sense. *Cogn. Dev.* 23, 409–422. doi: 10.1016/j.cogdev.2008.07.002
- Schneider, M., Merz, J., Sticker, J., De Smedt, B., Torbeyns, J., Verschaffel, L., et al. (2018a). Associations of number line estimation with mathematical competence: a meta-analysis. *Child Dev.* doi: 10.1111/cdev.13068 [Epub ahead of print]. doi: 10.1111/cdev.13068
- Schneider, M., Thompson, C. A., and Rittle-Johnson, B. (2018b). “Associations of magnitude comparison and number line estimation with mathematical competence: a comparative review,” in *Cognitive Development from a Strategy Perspective: A Festschrift for Robert S. Siegler*, ed. P. Lemaire (London: Psychology Press), 100–119.
- Shanghai Municipal Education Commission (2004). *Shanghai Primary and Secondary Mathematics Curriculum Standard (Trial Version)*. Shanghai: Shanghai Education Publisher.
- Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x
- Siegler, R. S., and Braithwaite, D. W. (2017). Numerical development. *Annu. Rev. Psychol.* 68, 187–213. doi: 10.1146/annurev-psych-010416-044101
- Siegler, R. S., and Lortie-Forgues, H. (2014). An integrative theory of numerical development. *Child Dev. Perspect.* 8, 144–150. doi: 10.1111/cdep.12077
- Siegler, R. S., and Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychol. Sci.* 19, 759–763. doi: 10.1111/j.1467-9280.2008.02153.x
- Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–250. doi: 10.1111/1467-9280.02438
- Smets, K., Sasanguie, D., Szűcs, D., and Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *J. Cogn. Psychol.* 27, 310–325. doi: 10.1080/20445911.2014.996568

- Starr, A., DeWind, N. K., and Brannon, E. M. (2017). The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. *Cognition* 168, 222–233. doi: 10.1016/j.cognition.2017.07.004
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *J. Exp. Psychol.* 18, 643–662. doi: 10.1037/h0054651
- Swanson, H. L. (2006). Cross-sectional and incremental changes in working memory and mathematical problem solving. *J. Educ. Psychol.* 98, 265–281. doi: 10.1037/0022-0663.98.2.265
- The National Bureau of Statistics in Shanghai (2017). *2017 Shanghai Statistical Yearbook*. Beijing: China Statistics Press.
- Tosto, M. G., Petrill, S. A., Malykh, S., Malki, K., Haworth, C., Mazzocco, M. M., et al. (2017). Number sense and mathematics: which, when and how? *Dev. Psychol.* 53, 1924–1939. doi: 10.1037/dev0000331
- van Marle, K., Chu, F. W., Li, Y., and Geary, D. C. (2014). Acuity of the approximate number system and preschoolers' quantitative development. *Dev. Sci.* 17, 492–505. doi: 10.1111/desc.12143
- Wang, L., Sun, Y., and Zhou, X. (2016). Relation between approximate number system acuity and mathematical achievement: the influence of fluency. *Front. Psychol.* 7:1966. doi: 10.3389/fpsyg.2016.01966
- Wechsler, D. (2009). *Wechsler Memory Scale: WMS-IV Technical and Interpretive Manual*. San Antonio, TX: Pearson.
- Wong, T. T. Y., Ho, C. S. H., and Tang, J. (2016). The relation between ANS and symbolic arithmetic skills: the mediating role of number-numerosity mappings. *Contemp. Educ. Psychol.* 46, 208–217. doi: 10.1016/j.cedpsych.2016.06.003
- Xenidou-Dervou, I., De Smedt, B., van der Schoot, M., and van Lieshout, E. C. (2013). Individual differences in kindergarten math achievement: the integrative roles of approximation skills and working memory. *Learn. Individ. Dif.* 28, 119–129. doi: 10.1016/j.lindif.2013.09.012
- Xenidou-Dervou, I., Molenaar, D., Ansari, D., van der Schoot, M., and van Lieshout, E. C. (2016). Nonsymbolic and symbolic magnitude comparison skills as longitudinal predictors of mathematical achievement. *Learn. Instr.* 50, 1–13. doi: 10.1016/j.learninstruc.2016.11.001
- Zhang, Y., Chen, C., Liu, H., Cui, J., and Zhou, X. (2016). Both non-symbolic and symbolic quantity processing are important for arithmetical computation but not for mathematical reasoning. *J. Cogn. Psychol.* 28, 807–824. doi: 10.1080/20445911.2016.1205074
- Zhu, M., Cai, D., and Leung, A. W. (2017). Number line estimation predicts mathematical skills: difference in Grades 2 and 4. *Front. Psychol.* 8:1576. doi: 10.3389/fpsyg.2017.01576

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Cai, Zhang, Li, Wei and Georgiou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Implications of Change/Stability Patterns in Children's Non-symbolic and Symbolic Magnitude Judgment Abilities Over One Year: A Latent Transition Analysis

Cindy S. Chew, Jason D. Forte and Robert A. Reeve*

Melbourne School of Psychological Sciences, University of Melbourne, Parkville, VIC, Australia

OPEN ACCESS

Edited by:

Marcus Lindskog,
Uppsala University, Sweden

Reviewed by:

Kenny Skagerlund,
Linköping University, Sweden
Maciej Haman,
University of Warsaw, Poland

*Correspondence:

Robert A. Reeve
r.reeve@unimelb.edu.au

Specialty section:

This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

Received: 15 May 2018

Accepted: 13 February 2019

Published: 05 March 2019

Citation:

Chew CS, Forte JD and Reeve RA
(2019) Implications
of Change/Stability Patterns in
Children's Non-symbolic
and Symbolic Magnitude Judgment
Abilities Over One Year: A Latent
Transition Analysis.
Front. Psychol. 10:441.
doi: 10.3389/fpsyg.2019.00441

Non-symbolic magnitude abilities are often claimed to support the acquisition of symbolic magnitude abilities, which, in turn, are claimed to support emerging math abilities. However, not all studies find links between non-symbolic and symbolic magnitude abilities, or between them and math ability. To investigate possible reasons for these different findings, recent research has analyzed differences in non-symbolic/symbolic magnitude abilities using latent class modeling and has identified four different magnitude ability profiles residing within the general magnitude ability distribution that were differentially related to cognitive and math abilities. These findings may help explain the different patterns of findings observed in previous research. To further investigate this possibility, we (1) attempted to replicate earlier findings, (2) determine whether magnitude ability profiles remained stable or changed over 1 year; and (3) assessed the degree to which stability/change in profiles were related to cognitive and math abilities. We used latent transition analysis to investigate stability/changes in non-symbolic and symbolic magnitude abilities of 109 5- to 6-year olds twice in 1 year. At Time 1 and 2, non-symbolic and symbolic magnitude abilities, number transcoding and single-digit addition abilities were assessed. Visuospatial working memory (VSWM), naming numbers, non-verbal IQ, basic RT was also assessed at Time 1. Analysis showed stability in one profile and changes in the three others over 1 year. VSWM and naming numbers predicted profile membership at Time 1 and 2, and profile membership predicted math abilities at both time points. The findings confirm the existence of four different non-symbolic–symbolic magnitude ability profiles; we suggest the changes over time in them potentially reflect deficit, delay, and normal math developmental pathways.

Keywords: non-symbolic and symbolic magnitude ability profiles, stability and change patterns, longitudinal analysis, visuospatial working memory, naming number ability, latent transition analysis

INTRODUCTION

Magnitude representation ability is as an important component of children's math ability (Siegler, 2016). Near-identical error and RT response signatures for non-symbolic magnitude judgments and symbolic magnitude judgments is claimed to reflect a common underlying representation – the approximate number system (ANS) where magnitudes are ordered akin to a mental number line (Moyer and Landauer, 1967; Feigenson et al., 2004; Gebuis et al., 2009; Izard et al., 2009; Piazza, 2010). Some claim that non-symbolic magnitude abilities scaffold the acquisition of symbolic (Arabic number) magnitude abilities, which, in turn, support the acquisition of math ability (Dehaene, 2007, 2011; Piazza and Izard, 2009; Piazza, 2010; Siegler, 2016). Others, in contrast, claim that non-symbolic and symbolic magnitude abilities are independent of each other and exert independent effects on emerging math abilities (De Smedt et al., 2009; Holloway and Ansari, 2009; Maloney et al., 2010; Sasanguie et al., 2012a). The fact that research can be cited in support of both claims implies the developmental significance of the relationship between non-symbolic and symbolic magnitude representation and children's math abilities is uncertain.

We suggest this uncertainty may be resolved by examining the relationship between patterns of differences in children's non-symbolic and symbolic magnitude representation abilities and their associated math and cognitive abilities over time. Given math ability likely depends on both general/number-specific abilities (Jordan et al., 2013; Träff, 2013); it is important to model different general/number-specific relationships with different magnitude representation profiles. We further suggest that such an examination may reveal information about potentially different magnitude representation developmental pathways distinguishing between typical and atypical pathways that underpin different math outcomes (Reeve et al., 2018).

Findings from longitudinal research examining the relationships between non-symbolic, symbolic magnitude judgment and math abilities over time are mixed (Sasanguie et al., 2012a,b; Kolkman et al., 2013; Xenidou-Dervou et al., 2016). Desoete et al. (2012), for instance, assessed 5- to 6-year-olds on three occasions and found no correlation between non-symbolic and symbolic judgment accuracy. However, children's non-symbolic and symbolic magnitude judgments were independently associated with math abilities. And 5- to 6-year-olds' non-symbolic judgments predicted their calculation ability 1 year later and arithmetic fact retrieval 2 years later. Further, symbolic judgments were also associated with calculation. Vanbinst et al. (2015b) also found non-symbolic and symbolic magnitude abilities independently predicted 6-year old's arithmetic accuracy and fact retrieval 1 year later; however, only symbolic magnitude ability predicted these outcomes 6 months later.

Others, in contrast, have found only symbolic magnitude abilities predict math abilities over time. Bartelet et al. (2014), for example, found 6-year-olds' symbolic judgment efficiency (accuracy/RT) predicted arithmetic achievement 1 year later, whereas non-symbolic judgment did not. Nonetheless, they found correlations between non-symbolic and symbolic

judgment, RT and efficiency measures. Similarly, Sasanguie et al. (2013) found 6- to 8-year olds' symbolic, but not non-symbolic, judgment speed correlated with timed arithmetic and a standardized math test 1 year later. However, they did not find a correlation between symbolic and non-symbolic judgments.

While methodological factors (e.g., magnitude judgment measures, sample size and age) may contribute to the aforementioned differences in findings (Price et al., 2012; Xenidou-Dervou et al., 2016), they fail to account for all differences (Chen and Li, 2014; Fazio et al., 2014; Chew et al., 2016; Schneider et al., 2016). We suggest that the variability in both cross-sectional and longitudinal developmental magnitude representation research findings may reflect the use of variable-oriented analytic approaches for analyzing magnitude ability data, which focuses on the relations between variables (e.g., using aggregated data in correlations and regression models).

Aggregate data methods tend to assume (1) homogeneity with respect to how variables of interest influence each other, (2) deviations from the mean reflect measurement error and (3) within-age variability is noise (see Chew et al., 2016 for a discussion). In terms of developmental changes, aggregate methods assume "universal" patterns of change where the focus is a general model of normative (average) developmental changes. These methods, however, may mask the presence of different patterns of magnitude abilities and, ipso facto, the possibility that different development models of magnitude representation development reside within a general data distribution (Chew et al., 2016). Aggregating data is a dubious practice when within-age variability is systematically related to patterns of inter-individual development (Dowker, 2008; Bouwmeester and Verkoeijen, 2012; Reeve et al., 2012; Gray and Reeve, 2016; Paul and Reeve, 2016). Insofar as different patterns of non-symbolic-symbolic magnitude ability relationships can be identified, they would not be represented by a general model that would comprise a summary of the mixture patterns (Siegler, 1987; Bergman et al., 2003; Collins and Lanza, 2013; Paul and Reeve, 2016).

Some researchers have argued for person-centered analytic approach to better understand the significance of individual differences in patterns of early math cognition (Dowker, 2008; Reeve et al., 2012, 2018; Chew et al., 2016). A person-centered approach (1) rejects the assumption that the entire population is homogeneous with respect to how variables influence each other, and (2) attempt to identify individuals characterized by different patterns of associations that are similar within subgroups but are different between subgroups (Laursen and Hoff, 2006).

Latent class analysis is a statistical model-based approach for partitioning heterogeneity in a population by identifying a small group of homogenous latent subgroups embedded within a set of measures (Lanza and Cooper, 2016). Individuals are assigned to the subgroup for which the posterior probability of belonging to that subgroup is the highest; calculated as a function of the observed data and parameter estimates (Vermunt and Magidson, 2013b; Lanza and Cooper, 2016). Latent profile analysis can be extended to model longitudinal data, where transitions over time in latent subgroup membership are also estimated in the model (i.e., latent transition analysis, LTA). While subgroup membership is assumed to be stable in latent

profile analysis (stable patterns of response characteristics), in LTA, individuals may change membership in latent profiles across time (see Hickendorff et al., 2018 for an analysis of the value of latent modeling for research on development and learning).

Chew et al. (2016) employed latent class analysis to determine whether different non-symbolic and symbolic magnitude (accuracy and judgment speed) ability profiles can be extracted from a general non-symbolic–symbolic magnitude ability distribution. They identified four different non-symbolic–symbolic magnitude ability profiles, three of which corresponded to the different pattern of findings identified in previous research (similarly good/bad non-symbolic–symbolic magnitude abilities; poor on symbolic relative to better non-symbolic magnitude abilities) (e.g., Halberda et al., 2008; Holloway and Ansari, 2009). These authors also found a previously unidentified fourth profile in which children displayed better symbolic magnitude ability relative to non-symbolic ability. Children assigned to this profile showed relatively superior symbolic magnitude judgment accuracy, albeit with longer response times. Moreover, the four identified magnitude abilities profiles were associated with different cognitive and math abilities. Chew et al. (2016) suggested that the different non-symbolic–symbolic magnitude/cognitive/math profiles reflect potentially different developmental patterns or models of math development. Children who possessed good or average non-symbolic and symbolic magnitude abilities showed relatively better visuospatial working memory, symbolic number access and math abilities. Children with poorer symbolic magnitude abilities, relative to non-symbolic abilities, performed poorer on a symbolic number access task and had poorer math abilities, compared to other magnitude profiles. Children in the fourth profile had relatively poorer visuospatial working memory and poorer math abilities. These findings highlight the fact that there is no single developmental model of magnitude representation underlying math abilities *per se*.

While this research highlight the value of latent profile analysis in potentially making sense of the heterogeneous distribution of non-symbolic and symbolic magnitude abilities and associated cognitive/math abilities in young children, the significance of their findings require explication in at least two ways. First, can Chew et al.'s (2016) findings be replicated? It has been argued that outcome of latent class modeling requires replication before claims can be made about the conceptual authenticity of identified profiles (Hickendorff et al., 2018). Second, since Chew et al.'s (2016) research was cross-sectional, we know little about the stability and/or change in the identified non-symbolic and symbolic magnitude ability profiles over time, or their relationship with cognitive and/or math abilities. The latter issue is particularly important. The degree to which deficits, delays or normal developmental profiles can be identified depends critically on longitudinal modeling (Reeve et al., 2012, 2018; Hickendorff et al., 2018). Nevertheless, both issues require answers before strong claims can be made about the developmental significance of different non-symbolic–symbolic magnitude ability profiles, especially with respect to the existence and significance of different magnitude representation developmental pathways.

The Current Study

We employed latent class modeling of children's non-symbolic and symbolic judgment responses, as well as of children's cognitive and math abilities, to investigate the significance of the stability and/or change in different patterns of magnitude representation longitudinally. Our aim was to better understand the nature and significance of individual differences in patterns of math development which may be reflected as typical and atypical pathways (Dowker, 2008; Reeve et al., 2018). We used LTA to investigate 5- to 6-year-olds' non-symbolic and symbolic magnitude judgment accuracy and RT signature patterns twice in 1 year. Our analytic approach is similar to the LTA modeling used by Reeve et al. (2018) who identified three distinct computation development trajectories, reflecting typical, delayed, and deficit pathways.

We assessed children's VSWM and symbolic access ability since these abilities are often associated with magnitude representation and math abilities (De Smedt and Gilmore, 2011; Friso-Van Den Bos et al., 2013; Vanbinst et al., 2015b; Paul and Reeve, 2016). VSWM is thought to support numerical magnitude processing, predicated on the proposition that magnitude information is spatially organized (Dehaene, 1992; Dehaene and Cohen, 1997; Zorzi et al., 2002; Dehaene et al., 2003; de Hevia et al., 2008). The speed and accuracy naming numbers (Arabic digits) has been used to assess the ability to access number symbols information (i.e., symbolic number knowledge) which is often invoked as an explanation for differences in symbolic magnitude abilities and in turn, math abilities (Rousselle and Noël, 2007; Berteletti et al., 2010; De Smedt and Gilmore, 2011). Naming number ability is also a marker of symbolic access difficulty (Chew et al., 2016). We also included basic RT and a general intelligence measure since math ability is often associated with them (Kytälä and Lehto, 2008; Geary, 2011; Luwel et al., 2013; Vanbinst et al., 2015b).

We assessed children's single-digit addition and transcoding abilities to evaluate the relationship between profile membership and math abilities. We examined single-digit addition and transcoding ("reading" number strings) since they are considered important for later math abilities (Geary, 2000; OECD, 2012; Vanbinst et al., 2015a). In Australia single-digit addition is introduced to children from kindergarten onward and often used as an outcome measure (see Paul et al., 2018).

Based on Chew et al.'s (2016) findings, we anticipated identifying a profile that exhibited good, and one exhibiting average, non-symbolic and symbolic magnitude abilities (i.e., similar non-symbolic–symbolic magnitude abilities), and a profile that possessed better non-symbolic relative to symbolic magnitude ability. We also expected to identify a profile that exhibited better symbolic relative to non-symbolic magnitude ability. We expected children assigned to a good non-symbolic–symbolic magnitude ability profile to reflect a typical change pathway, and would exhibit good VSWM and naming number ability, and in turn, good single-digit addition and transcoding abilities across time. Insofar as other profiles reflect atypical change pathways, we expect children assigned to better non-symbolic relative to symbolic magnitude ability profile to possess poorer VSWM and those assigned to better symbolic relative

to non-symbolic magnitude ability profile to possess poorer naming number ability. Children displaying relatively poorer non-symbolic and/or poorer symbolic magnitude abilities would also perform poorer on single-digit addition and transcoding. While some children may move from one profile to another over time, we do not expect a child who belonged to a better performing profile (relative to other children) would move to a poorer profile over time.

MATERIALS AND METHODS

Participants

One-hundred-nine children (55% females) participated, comprising 48 Kindergarten ($M = 5.8$ years, $SD = 2.8$ months) and 61 Year 1 ($M = 6.8$ years, $SD = 3.6$ months) children at initial assessment. All spoke English, had normal or corrected-to-normal vision and had no identified learning difficulties. The study was conducted with the approval of, and in accordance with, the authors' University's human research ethics committee. The parents of children provided informed consent for their children to participate in the study.

Procedure

All children individually completed non-symbolic and symbolic magnitude judgments, naming numbers, single-digit addition, reading numbers, Corsi Blocks Backward (VSWM), Raven's Colored Progressive Matrices (non-verbal IQ) and basic RT tasks on the first assessment. Approximately 1 year later, they completed the non-symbolic and symbolic judgment tasks, and the single-digit addition and reading numbers tasks. Tasks were completed in short sessions over 3 days (non-symbolic and symbolic tasks were completed on separate days to avoid inter-task priming effects). Except for the non-verbal IQ and VSWM tasks, stimuli were presented on a 15" screen laptop computer running E-Prime software (version 2.0). The screen was at eye-level, approximately 40 cm in front of children. A fixation cross appeared in the center of the screen prior to a target stimulus appearing. Except for non-symbolic and symbolic tasks, in which response time was capped at 5,000 ms, stimuli remained on the screen until a response was made.

Non-symbolic and Symbolic Judgment Tasks

In the non-symbolic judgment task, two sets of blue squares separated by a central vertical line appeared on the screen (Chew et al., 2016). Children selected the set that had the most squares by pressing the corresponding right shift key or the left shift key. The task comprised 72 trials with judgment combinations of all quantities between one and nine blue squares, except ties (e.g., 9 and 9). The ratios for each trial (i.e., smaller number/larger number) were divided into eight ratios: 0.1–0.19; 0.2–0.29... up to 0.8–0.89. Stimuli were presented in a fixed random order, with the larger set appearing on the left- and right-hand sides of the screen equally. To reduce possible reliance on perceptual cues for judgments, individual square sizes and total area were

systematically varied across trials (total area was the same for both sets within trials) (Dehaene et al., 2005). We analyzed two indices (mean accuracy and median RT) (Bartelet et al., 2014; Ratcliff et al., 2015; Schneider et al., 2016), both of which are associated with math ability (De Smedt et al., 2013).

The symbolic judgment task was procedurally identical to the non-symbolic task, except black Arabic digits were presented on white background (60-point font size).

Number Naming

Children named digits between 1 and 9; each digit was presented three times in separate blocks of trials ($n = 27$ trials overall). The interviewer pressed a response key following each response and recorded responses verbatim (the interviewer could not see the computer screen). Median RT was used for analysis since children made few errors.

Single-Digit Addition

Children completed 30 two-term addition problems, following two practice trials. They were instructed to answer problems as quickly and as accurately as possible. Addends comprised combinations of all digits between "2" and "7" (excluding tied pairs: e.g., $2 + 2$), in both orders (e.g., $2 + 7$ and $7 + 2$). Single-digit addition problems are widely used as a measure of early computation ability (Bailey et al., 2012; Paul and Reeve, 2016).

Transcoding: Reading Multi-Digit Numbers

Children read 30 two to four digit numbers displayed on the computer screen (i.e., 11, 12, 14, 16, 17, 19, 28, 35, 47, 52, 73, 94, 105, 162, 207, 435, 574, 809, 1002, 2584, 3201, 4783, 6057, 9236, 10006, 26103, 50316, 46927, 60935, and 79768). The numbers were presented in the same randomized order for all children.

Corsi Blocks Backward (VSWM)

The interviewer tapped a sequence of blocks in a pre-specified order and children attempt to repeat the tap sequences in reverse order (Kessels et al., 2000). Children were ensured that they understood the task in preceding practice trials. Testing ceased after two failed trials. The VSWM span comprised the average of the longest correct reverse block tap sequences.

Raven's Colored Progressive Matrices (Non-verbal IQ)

RCPM was administered following manual instructions and responses scored using published age norms (Raven et al., 1995; Cotton et al., 2005).

Basic RT

The task comprised nine trials. Children pressed a computer key as quickly as possible when a black dot appeared on the screen approximately 500 ms later after a central fixation point.

Analytic Approach

We used LTA to identify distinct profiles (i.e., subgroups) of children who share similar non-symbolic-symbolic magnitude

judgment accuracy/RT response patterns, and examined changes in profile membership over time (Latent GOLD 5.1; Vermunt and Magidson, 2015). Similar to latent profile analysis, we rely on a set of criteria for selecting the optimal model solution (Trezise and Reeve, 2014; Chew et al., 2016). Goodness-of-fit statistics (e.g., Bayesian information criterion) weigh the fit of the models relative to the number of parameters, with a lower value indicating a better fitting model to the data (Vermunt and Magidson, 2013a). Entropy, which range from 0 to 1, assess how well the subgroups are classified and values greater than 0.8 are considered to have high entropy which implies better classification (Clark and Muthen, 2009). The theoretical relevance and usefulness of the latent profiles were also considered (Muthen and Muthen, 2000). Models were fit using 200 random starting sets and 500 replications to ensure that model convergence could be replicated.

The LTA model includes three types of parameters. It yields the conditional response probabilities that describe response patterns conditional on latent subgroup membership. For example, a profile with a relatively high probability of high accuracy and RT on non-symbolic/symbolic judgments can be interpreted as showing good non-symbolic–symbolic magnitude abilities. The model also yields class probabilities, which describe the size of each latent subgroup at each time point (i.e., relative frequency of class membership) and a matrix of transition probabilities (i.e., conditional probabilities describing the probability of being in a given subgroup at time = t , conditional on the subgroup at time = $t - 1$) which describes how children transition from Time 1 to Time 2 in non-symbolic–symbolic magnitude ability profiles. Measurement invariance was modeled (i.e., conditional response probabilities are the same across the two time points), following from previous work (Chew et al., 2016) and initial examination showing consistency in profiles at both time points. The same number and type of classes occur at both time points allowing a straightforward interpretation since the meanings of the profiles are the same across time.

The following covariates were included in the model as predictors of latent profiles at Time 1 and 2, as well as predictors of transitions in profile membership between Time 1 and 2: VSWM, naming numbers, basic RT, non-verbal IQ and grade. When covariates are included in the LTA model (i.e., in a 1-step model), current profile membership (i.e., described by transitional probabilities) is predicted by both profile membership at the previous time point and the value of the covariates. Class profiles, class sizes and transition probabilities may change as a result.

A three-step estimation procedure (Vermunt and Magidson, 2015) was separately conducted for the LTA model where the association between the predictor variables and assigned membership are examined at time points and the underlying statistical model is analogous to a multinomial regression logistic regression. The step-three modeling approach allows for the correction of classification errors obtained when assigning profile memberships (maximum-likelihood adjustment method is used to correct for classification errors) at the particular time points—a failure to account for classification errors can lead to an underestimation of the relationship between profile membership

and other variables (Bakk et al., 2013). This estimation approach is desirable in the LTA context because the 1-step model approach (i.e., covariates included in model) has the drawback that covariate values at one point in time affects the definition of the latent class variable at another point in time. Similarly, SDA and transcoding abilities (treated here as dependent variables) were regressed on the latent profile membership at Time 1 and 2.

RESULTS

As expected, non-symbolic and symbolic magnitude RT and error rates increased with increasing ratios and decreased with increasing grade (descriptive statistics are reported in **Supplementary Material**). Means and standard deviations for all measures as a function of grade are reported in **Table 1**. Zero-order correlations among measures are reported in **Table 2** which shows significant correlations between children's non-symbolic–symbolic magnitude judgments, and SDA problem solving and transcoding abilities. Non-symbolic and symbolic accuracy/RTs were correlated at Time 1 and Time 2. Similarly, non-symbolic and symbolic accuracy and RT at Time 1 correlated with the same measures at Time 2 (except non-symbolic RT).

Assessing Model Fit

Models comprising one to six latent profiles were estimated from non-symbolic and symbolic magnitude RT and accuracy at Time 1 and 2. Goodness-of-fit indices for each model are reported in **Table 3**: a four-profile solution was selected as the best fitting model. A three and five profile solution were also considered; the three-profile solution was not optimal when analyzed at Time 1 and 2 separately while the five profile had low interpretability (e.g., the fifth profile had a small number of children who appeared similar to another profile). The four-profile model was selected on the basis of fit, previous research (Chew et al., 2016) and conceptual interpretability. Model selection was supported by a high entropy value (i.e., above 0.8), indicating good classification of individuals into latent profiles.

Non-symbolic–Symbolic Magnitude Ability Profiles

Deviations from overall mean proportion accuracy (non-symbolic = 0.88; symbolic = 0.89) and median RT (non-symbolic = 1236.54 ms; symbolic = 1179.84 ms) for the four profiles across Time 1 and 2 are presented in **Figure 1**. Labels corresponding to the relative non-symbolic and symbolic magnitude abilities were assigned to each profile. (Note, the numbers attached to the profiles—i.e., profile 1, 2, 3, 4—are convenient labels and not a statement about the ordinal position of the profiles.) Profile 1 comprised children who displayed average accuracy and speed on both non-symbolic and symbolic judgments (i.e., relatively close to the overall mean/median). They were characterized by “average non-symbolic–symbolic magnitude abilities.” Profile 2 included children who were relatively highly accurate and fast on both non-symbolic and symbolic judgments, and they were characterized by

TABLE 1 | NSM and SM measures, cognitive factors, and math abilities as a function of grade.

	Grade			
	Kindergarten	Year 1	Year 1	Year 2
	<i>M (SD) at Time 1</i>		<i>M (SD) at Time 2</i>	
NSM correct	0.79 (0.14)	0.87 (0.08)	0.92 (0.06)	0.94 (0.05)
NSM RT	1479.53 (387.84)	1287.22 (317.95)	1092.89 (231.06)	958.08 (271.72)
SM correct	0.82 (0.16)	0.89 (0.08)	0.91 (0.05)	0.93 (0.05)
SM RT	1344.43 (416.02)	1183.39 (273.18)	1148.64 (199.7)	905.47 (255.1)
Transcoding	42.53 (23.87)	80.94 (15.09)	75 (18.79)	96.65 (6.36)
SDA correct	42.57 (35.78)	91.04 (9.9)	86.67 (19.51)	95.03 (8.25)
VSWM	3.25 (0.64)	3.6 (0.85)	—	—
Naming Numbers	956.56 (242.05)	840.97 (190.59)	—	—
RCPM	80.73 (22)	83.52 (18.33)	—	—
Basic RT	632.70 (82.46)	589.47 (88.47)	—	—

NSM, non-symbolic magnitude; SM, symbolic magnitude; M, means except median for NSM RT, SM RT, naming numbers, and basic RT.

TABLE 2 | Zero-order correlations among NSM-SM measures, cognitive factors, and math abilities.

	Time 1				Time 2				Time 1				Time 2			
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Time 1	1															
1. NSM correct	1															
2. SM correct	0.44**	1														
3. NSM RT	−0.26**	−0.02	1													
4. SM RT	−0.18	0.04	0.54**	1												
Time 2					1											
5. NSM correct	0.29**	0.28**	−0.17	−0.1	1											
6. SM correct	0.2*	0.45**	−0.06	0.04	0.31**	1										
7. NSM RT	−0.3**	−0.23*	0.17	0.13	−0.41**	−0.13	1									
8. SM RT	−0.2*	−0.37**	0.24*	0.3**	−0.34**	−0.24*	0.64**	1								
9. VSWM	0.14	0.35**	−0.11	−0.09	0.26**	0.18	−0.2*	−0.2*	1							
10. NN	−0.11	−0.18	0.18	0.08	−0.24*	−0.12	0.4**	0.4**	−0.04	1						
11. RCPM	0.1	0.16	0.04	−0.02	0.15	0.13	−0.09	−0.18	0.19*	−0.16	1					
12. Basic RT	−0.26**	−0.05	0.24*	0.28**	−0.06	0.08	0.13	0.14	−0.1	0.21*	0.06	1				
Time 1													1			
13. SDA correct	0.31**	0.42**	−0.21*	−0.15	0.17	0.19*	−0.31**	−0.33**	0.35**	−0.33**	0.16	−0.16	1			
14. Transcoding	0.21*	0.42**	−0.12	−0.2*	0.19	0.32**	−0.26**	−0.43**	0.28**	−0.43**	0.11	−0.18	0.72**	1		
Time 2																
15. SDA correct	0.22*	0.42**	0.03	−0.07	0.31**	0.4**	−0.33**	−0.33**	0.25**	−0.34**	0.25**	−0.11	0.28**	0.42**	1	
16. Transcoding	0.37**	0.5**	−0.14	−0.2*	0.22*	0.33**	−0.38**	−0.47**	0.31**	−0.42**	0.19*	−0.2*	0.68**	0.73**	0.49**	1

NSM, non-symbolic magnitude; SM, symbolic magnitude; NN, naming numbers RT. * $p < 0.05$, ** $p < 0.01$.

“good non-symbolic–symbolic magnitude abilities.” Children in Profile 3 were relatively average on symbolic judgments but much less accurate on non-symbolic judgments relative to non-symbolic. They also exhibited relatively long response speed on both non-symbolic and symbolic judgments. Hence, they were characterized by “better symbolic abilities relative to non-symbolic abilities.” Children in Profile 4 were less accurate and slower on both non-symbolic and symbolic judgments relative to other children. However, they were more accurate on non-symbolic relative to symbolic judgments. They were characterized by “better non-symbolic abilities relative to symbolic abilities.” One-way ANOVAs and Bonferroni-corrected *post hoc* comparisons showed profiles differed from each other in non-symbolic and symbolic magnitude, accuracy and RT (details are reported in the **Supplementary Material**).

Change/Stability Patterns in Non-symbolic–Symbolic Magnitude Profiles Over Time

Table 4 presents the transition probabilities, which reflect the probability of a child transitioning to a particular profile at Time 2, conditional on their profile membership at Time 1. These parameters describe the patterns of change in non-symbolic–symbolic magnitude abilities across time. (The probabilities may also be considered to reflect the proportion of each profile at Time 1 that transitioned into particular profiles at Time 2.) Diagonal values indicate the proportion of children who remained in the same profile at both times. Off-diagonal values indicate the proportion of children in a particular profile at Time 1, who move into another profile at Time 2. Results indicate that membership

TABLE 3 | Fit information for the latent transition analysis model.

N Profiles	N par	Log			Entropy	
		likelihood	aBIC	AIC3	CAIC	R-squared
1	8	−2781.95	5581.62	5587.89	5609.42	1
2	19	−2536.91	5115.91	5130.81	5181.95	0.80
3	32	−2453.60	4978.08	5003.18	5089.31	0.83
4	47	−2420.46	4945.05	4981.92	5108.41	0.83
5	64	−2396.88	4935.55	4985.75	5158.00	0.82
6	83	−2380.38	4944.65	5009.75	5233.14	0.85

Bold values indicate best fitting model. N Profiles, number of latent profiles; N par, number of parameters in the model; aBIC, adjusted Bayesian information criteria; AIC3, Akaike's information criterion with 3 as penalizing factor; CAIC, consistent AIC.

to Profile 2 was stable; 96% ($n = 4$) of children who were in Profile 2 at Time 1, remained in Profile 2 at Time 2. In other words, it is very unlikely (low probability) that children in this Profile would move into any of the other Profiles.

Children who were in Profile 1 at Time 1 had a high likelihood (0.97 probability) of moving into Profile 2 at Time 2. That is, 97% ($n = 38$) of the Profile 1 children at Time 1 moved into Profile 2 at Time 2. It was unlikely that children in this Profile moved into other profiles. Sixty-seven percent of children ($n = 31$) who were in Profile 3 at Time 1 moved into Profile 1 at Time 2; followed by Profile 2 (29%; $n = 13$). Finally, of the children in Profile 4 at Time 1, 48% ($n = 10$) moved into Profile 1 and 42% ($n = 8$) into Profile 3 at Time 2. It was rare (10%; $n = 2$) that they moved into Profile 2 at Time 2. Of note, all children in Profile 4 at Time 1 moved into other profiles at Time 2 and no children moved into this profile at Time 2 (i.e., there were no children in Profile 4 at Time 2).

In sum, most children change in profile membership over time, except those in Profile 2 who exhibited stability. Specifically, children moved to a better non-symbolic–symbolic magnitude ability profile over time. Frequencies in profiles as a function of grade at Time 1 and 2 are presented in Table 5.

TABLE 4 | Latent transition probabilities based on latent transitional analysis model for NSM and SM.

Time 1 status	Time 2 status				
	Profile 1	Profile 2	Profile 3	Profile 4	N
Profile 1	0.0196	0.9727	0.0059	0.0017	39
Profile 2	0.0138	0.9578	0.0144	0.0139	4
Profile 3	0.6745	0.2898	0.0342	0.0015	46
Profile 4	0.4770	0.1041	0.4153	0.0036	20
N	42	57	10	0	109

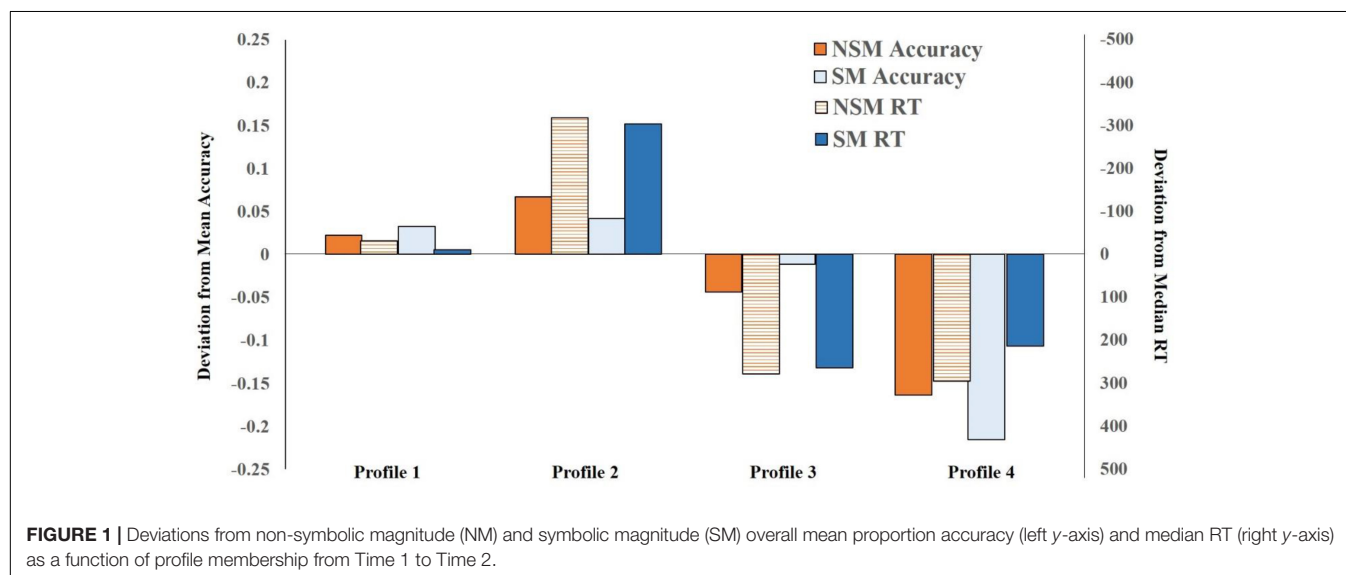
Bold values indicate the proportion of children who remained in the same profile at T1 and T2. NSM, non-symbolic magnitude; SM, symbolic magnitude.

TABLE 5 | Frequencies in non-symbolic–symbolic magnitude profiles as a function of grade at Time 1 versus Time 2.

	Time 1		Time 2	
	Kindergarten	Year 1	Year 1	Year 2
Profile 1	8	31	26	16
Profile 2	0	4	15	42
Profile 3	24	22	7	3
Profile 4	16	4	0	0

Predicting Non-symbolic–Symbolic Magnitude Ability Profiles

To examine whether cognitive measures predicted transitions in non-symbolic–symbolic magnitude profiles from Time 1 to 2, five predictors were included in the LTA model. (Grade was included to examine possible age-related effects on profile membership across time.) The overall model showed VSWM ($Wald = 6.56$, $p = 0.88$), non-verbal IQ ($Wald = 7.16$, $p = 0.85$), basic RT ($Wald = 9.22$, $p = 0.69$), naming number ability ($Wald = 6.77$, $p = 0.87$) and grade ($Wald = 4.69$, $p = 0.97$) did not reach statistical significance. Next, we used the three-step procedure



to determine whether initial cognitive measures/age (Time 1) predicted profile membership at both time points. We examined the standardized regression coefficients (*z*-scores) and Wald statistics for each measure predicting profile memberships in a multivariate model that accounts for classification errors (see **Table 6**). The *z*-scores show the predictive effect of factors for profiles while taking into account other variables in the model. Findings show VSWM and naming numbers independently predict non-symbolic–symbolic magnitude profiles at both Time 1 and 2, whereas non-verbal IQ and basic RT did not. Age was only associated with profile membership at Time 1.

At Time 1, an increase in age was associated with an increased likelihood of belonging to Profile 2 ($B = 0.24$, $SE = 0.11$, $z = 2.11$) and a reduced likelihood of belonging to Profiles 3 ($B = -0.1$, $SE = 0.04$, $z = -2.18$) and 4 ($B = -0.15$, $SE = 0.06$, $z = -2.76$). An increase in VSWM was associated with an increased likelihood of belonging to Profile 2 ($B = 2.25$, $SE = 1.09$, $z = 2.06$) and conversely, a reduced likelihood of belonging to Profile 4 ($B = -1.8$, $SE = 0.56$, $z = -3.23$). A poorer naming number ability (i.e., longer naming number RT) was associated with a reduced likelihood of belonging to Profile 2 ($B = -0.01$, $SE = 0.004$, $z = -2.74$) and a greater likelihood of belonging to Profiles 3 ($B = 0.004$, $SE = 0.002$, $z = 2.44$) and 4 ($B = 0.005$, $SE = 0.002$, $z = 3.09$).

At Time 2, an increase in VSWM was associated with a greater likelihood of belonging to Profile 2 ($B = 0.74$, $SE = 0.3$, $z = 2.48$). A poorer naming number ability was associated with a reduced likelihood of belonging to Profile 2 ($B = -0.004$, $SE = 0.001$, $z = -3.45$) and a greater likelihood of belonging to Profile 3 ($B = 0.004$, $SE = 0.001$, $z = 2.96$).

Overall, at Time 1, children in Profile 2 were more likely to be older, and conversely children in Profiles 3 and 4 were more likely to be younger. However, at Time 2, age was no

longer associated with profile membership. At Time 1, higher VSWM and naming number ability characterized children in Profile 2, whereas poorer VSWM and naming number ability characterized children in Profile 4. Poorer naming number ability also characterized children in Profile 3. At Time 2, higher VSWM and naming number ability remained characteristic of children in Profile 2, whereas poorer naming number ability remained characteristic of children in Profile 3.

Non-symbolic–Symbolic Magnitude Profiles Predicting Math Abilities

SDA and transcoding were regressed on profile membership at Time 1 and 2 while accounting for classification errors using the three-step procedure. Accuracy reading teen, two digit, three digit and four digit numbers showed reasonably good internal consistency (Cronbach alpha = 0.77) and hence, accuracy was summed across these digit strings. The standardized regression coefficients (*z*-scores) and Wald statistics for each dependent variable predicted by profile membership for both time points are reported in **Table 7**. SDA correctness and transcoding were significantly associated with profile membership at Time 1 and 2. At Time 1, an increase in SDA accuracy was associated with belonging to Profiles 1 and 2 and, conversely, a reduced likelihood of belonging to Profile 4. An increase in transcoding ability was associated with belonging to Profiles 1 and 2 and, conversely, a reduced likelihood of belonging to Profiles 3 and 4. At Time 2, an increase in SDA accuracy and transcoding was associated with belonging to Profile 2.

DISCUSSION

The purpose of the present study was to assess the degree to which (1) the different non-symbolic–symbolic magnitude representation profiles identified in a previous study would be re-identified, and (2) profiles remained stable or changed over time. The aim was to determine whether stability/change in profiles were related to children's cognitive and math abilities. Of interest was whether we could identify different magnitude representation pathways that distinguish typical and atypical models of math development. Four findings are of note. First, the current study replicates Chew et al. (2016) by showing that four meaningfully different non-symbolic–symbolic magnitude ability profiles can be extracted from a general non-symbolic–symbolic magnitude ability distribution. Second, the change/stability in profiles across time suggests different magnitude representation developmental pathways can be identified. Third, VSWM and naming number abilities were associated with profile membership at Time 1 and Time 2; however, they did not predict stability/change in profile membership over time. Fourth, non-symbolic–symbolic magnitude ability profiles were differentially associated with math abilities at Time 1 and Time 2.

Non-symbolic–Symbolic Magnitude Ability Profiles

The mean accuracy and median RTs for both non-symbolic and symbolic magnitude judgments of children in Profile 1 were close

TABLE 6 | Time 1 covariates predicting NSM-SM profile memberships at Time 1 and Time 2.

NSM-SM profiles						
<i>N</i> = 160	<i>Wald</i>	<i>P</i>	Profile 1 <i>z</i> -score	Profile 2 <i>z</i> -score	Profile 3 <i>z</i> -score	Profile 4 <i>z</i> -score
Time 1						
Age	11.4	0.009	0.33	2.11*	−2.18*	−2.76**
VSWM	10.68	0.014	−0.37	2.06*	−0.7	−3.23**
Naming number	9.72	0.021	0.94	−2.74**	2.44*	3.09**
RCPM	7.32	0.062	2.38	−2.21	0.053	1.25
Basic RT	6.05	0.11	2.01	−2.14	2.28	0.98
Time 2						
Age	6.81	0.033	−1.53	1.73	−0.12	−
VSWM	7.02	0.03	−0.26	2.48*	−1.32	−
Naming number	12.42	0.002	0.41	−3.45**	2.96**	−
RCPM	2.01	0.37	−1.07	1.04	0.15	−
Basic RT	0.34	0.84	0.52	0.53	−0.58	−

Bold values indicate significant predictors. NSM, non-symbolic magnitude; SM, symbolic magnitude. * $p < 0.05$, ** $p < 0.01$.

TABLE 7 | Significant effects in three-step latent profile model with dependent variables at Time 1 and Time 2.

		NSM–SM profiles				
<i>N</i> = 160	<i>Wald</i>	<i>p</i>	Profile 1	Profile 2	Profile 3	Profile 4
Time 1						
SDA	60.16	<0.001	<i>B</i> = 18.51, <i>SE</i> = 3.81, <i>z</i> = 4.86**	<i>B</i> = 22.09, <i>SE</i> = 3.49, <i>z</i> = 6.34**	<i>B</i> = −7.59, <i>SE</i> = 5.13, <i>z</i> = −1.48	<i>B</i> = −33.01, <i>SE</i> = 7.17, <i>z</i> = −4.60**
Transcoding	44.47	<0.001	<i>B</i> = 16.52, <i>SE</i> = 4.07, <i>z</i> = 4.06**	<i>B</i> = 14.54, <i>SE</i> = 5.25, <i>z</i> = 2.77*	<i>B</i> = −8.77, <i>SE</i> = 3.72, <i>z</i> = −2.36*	<i>B</i> = −22.29, <i>SE</i> = 4.89, <i>z</i> = −4.56**
Time 2						
SDA	13.62	0.001	<i>B</i> = 3.18, <i>SE</i> = 4.38, <i>z</i> = 0.73	<i>B</i> = 10.49, <i>SE</i> = 3.63, <i>z</i> = 2.89**	<i>B</i> = −13.67, <i>SE</i> = 7.52, <i>z</i> = −1.82	–
Transcoding	38.81	<0.001	<i>B</i> = −3.06, <i>SE</i> = 3.75, <i>z</i> = −0.82	<i>B</i> = 14.37, <i>SE</i> = 3.09, <i>z</i> = 4.64**	<i>B</i> = −11.31, <i>SE</i> = 6.06, <i>z</i> = −1.87	–

NSM, non-symbolic magnitude; SM, symbolic magnitude. **p* < 0.05, ***p* < 0.01.

to the average mean accuracy and median RTs for the entire sample. Children in Profile 2 were more accurate and faster making both non-symbolic and symbolic magnitude judgments, relative to children in other profiles. The non-symbolic–symbolic magnitude judgment response patterns of children in Profiles 1 and 2 are consistent with claims made for an association between non-symbolic and symbolic magnitude abilities (Piazza et al., 2010; Dehaene, 2011; Feigenson et al., 2013).

Children in Profile 3 were more accurate in their symbolic magnitude judgments, relative to their non-symbolic judgments, but compared to Profiles 1 and 2, they were also relatively slower in making non-symbolic and symbolic magnitude judgments. This pattern of non-symbolic–symbolic magnitude judgment replicates Chew et al.'s (2016) findings. They suggested that symbolic abilities can be supported by rote practice. Some children may learn by rote recall and complete the symbolic judgments with some success but doing so requires more effort (i.e., longer RT and hence, possibly less efficient).

Children in Profile 4 were less accurate and slower making non-symbolic and symbolic magnitude judgments, compared to children in the other three profiles. However, they were more accurate making non-symbolic judgments compared to symbolic judgments, which is consistent with claims that symbolic magnitude abilities can be independent of non-symbolic abilities (i.e., better non-symbolic abilities relative to symbolic abilities—see Rousselle and Noël, 2007; Holloway and Ansari, 2009; Sasanguie et al., 2014).

Change/Stability Patterns in Non-symbolic–Symbolic Magnitude Ability Profiles

Most children's non-symbolic and symbolic magnitude abilities changed across time. The general “movement” pattern was from a less accurate and slower non-symbolic–symbolic magnitude

profile to a more accurate and faster ability profile. No child moved from a better ability profile at Time 1 to a poorer one at Time 2. Only a small group of children (96%; *n* = 4) were stable across time (i.e., remained in Profile 2 across time); this stability suggests a consistency in good non-symbolic–symbolic magnitude abilities across time. Almost all children (97%; *n* = 38) moved from Profile 1 to Profile 2 at Time 2. While some children (29%; *n* = 13) moved from Profile 3 to Profile 2 at Time 2, the majority (67%; *n* = 31) moved to Profile 1. Similar proportions of children moved from Profile 4 to Profiles 1 (48%; *n* = 10) and 3 (42%; *n* = 8) at Time 2. Children in this profile rarely moved to Profile 2 at Time 2 (10%; *n* = 2).

The change in profile membership from Profile 1 (average non-symbolic–symbolic magnitude ability) to Profile 2 (good non-symbolic–symbolic magnitude ability) at Time 2 could be regarded as representing an expected change pathway. However, other changes in profile membership over time (e.g., Profile 4 to Profile 3) suggest that there may be other, possibly less optimal, change pathways to consider (i.e., atypical pathway). While there may be more than one non-symbolic–symbolic magnitude developmental pathway, they may represent different routes to competency (equifinality) or indicators of difficulties over time. For instance, the movement of some Profile 4 children (relatively better non-symbolic to symbolic magnitude ability) to Profile 3 (relatively better symbolic to non-symbolic magnitude ability) at Time 2 suggests that some children continue to develop symbolic abilities, separate from non-symbolic abilities. These children may represent relatively poorer developmental change (possibly reflecting a math delay or a deficit) in that they are not transitioning into a profile better on both symbolic and non-symbolic magnitude abilities. The movement of other Profile 4 children to Profile 1 at Time 2 suggests some children do continue to improve in both non-symbolic and symbolic magnitude abilities – consistent with claims that non-symbolic magnitude ability supports symbolic ability.

The question of multiple developmental routes to equifinality or even math difficulties cannot be answered with only two time points, 1 year apart. However, current findings caution against the assumption of one general developmental pathway. Since different non-symbolic–symbolic magnitude ability profiles exist, it would be inappropriate to represent these two magnitude representation ability by a general model that reflects normative developmental changes (i.e., variable-centered analytical approaches). Using LTA allowed us to examine individual differences in patterns of change over time in which more than one developmental trajectories can systematically differ across individuals.

Cognitive Factors/Age and Non-symbolic–Symbolic Magnitude Ability Profiles

While older children were likely to belong to Profile 2 and younger children were more likely to belong to Profiles 3 and 4; grade only partially overlapped with profile membership. Children from both grades were represented in all profiles at Time 1 (except no kindergartener children were assigned to Profile 2). At Time 2, age was not associated with profile membership. Using LTA to characterize age variability allowed us to sidestep the assumption of age as proxy for development and, examine how age is related to the magnitude profiles a posteriori and how the cognitive factors related to profiles after taking age into account. Our findings are consistent with recent studies that caution against focusing on age differences which may mask meaningful profiles of competence (Gray and Reeve, 2014; Paul and Reeve, 2016).

VSWM, naming number ability, non-verbal IQ, basic RT nor age predicted changes in non-symbolic–symbolic magnitude ability profile membership across time. However, VSWM and naming number abilities at Time 1 were associated with profile memberships at both time points. This finding is consistent with studies that have found a link between VSWM/naming number abilities and math ability in young children (De Smedt et al., 2009; Berteletti et al., 2010; Mammarella et al., 2010; Vanbinst et al., 2015b).

At Time 1, poorer VSWM was associated with Profile 4; and poorer naming number ability was associated with Profiles 3 and 4. Conversely, good VSWM and naming number ability were associated with Profile 2. Similarly, at Time 2, good VSWM and naming number ability were associated with Profile 2 while poorer naming number ability was associated with Profile 3. The findings are consistent with claims that the ability to access symbolic numerical information is a number-specific cognitive factor for children who show poor symbolic magnitude abilities (i.e., Profile 4) (Rousselle and Noël, 2007; Holloway and Ansari, 2009).

Of interest, poorer naming number ability is also characteristic of children who displayed better symbolic magnitude abilities relative to non-symbolic abilities (Profile 3) at both time points. While children in Profile 3 completed symbolic magnitude judgments accurately, they took longer in making judgments. Nevertheless, their basic RT was not significantly

different to other children. Naming numbers may be a useful marker of children's ability to efficiently access symbolic magnitude information (Vanbinst et al., 2015b). Indeed, good naming number ability predicted good non-symbolic–symbolic magnitude abilities (Profile 2) at first test occasion and 1 year later.

Magnitude information is argued to be encoded spatially and VSWM is implicated in the representation and manipulation of numerical magnitudes more generally (Zorzi et al., 2002; Dehaene and Brannon, 2011; de Hevia, 2016). Although poorer VSWM was associated with Profile 4 children (i.e., relatively better non-symbolic to symbolic magnitude abilities), they were also children who displayed the weakest non-symbolic abilities at initial assessment (all moved out of the profile at Time 2). On the other hand, good VSWM predicted good non-symbolic–symbolic magnitude abilities (i.e., Profile 2) at both time points. This finding suggest that VSWM capacity may underpin non-symbolic magnitude abilities and, in turn, symbolic magnitude development and math abilities for some children.

Non-symbolic–Symbolic Magnitude Ability Profiles and Math Abilities

Non-symbolic–symbolic magnitude ability profiles were associated with SDA and transcoding at both Time 1 and 2. This is consistent with research which has found a link between non-symbolic and/or symbolic magnitude abilities and math abilities (De Smedt et al., 2009; Bugden and Ansari, 2011; Mazzocco et al., 2011; Xenidou-Dervou et al., 2016). At Time 1, children with good/average non-symbolic–symbolic magnitude abilities relative to other children (i.e., Profiles 1 and 2) were associated with better SDA problem-solving and transcoding. In contrast, children with better non-symbolic abilities relative to symbolic abilities (Profile 4), was associated with poorer SDA problem-solving and transcoding. This finding is consistent with studies that show poorer symbolic magnitude judgment (and symbolic number access) are associated with poorer math abilities (Rousselle and Noël, 2007; De Smedt and Gilmore, 2011).

Children with better symbolic abilities, relative to non-symbolic abilities (Profile 3), possessed poorer transcoding skills—a finding that runs contrary to claims that symbolic magnitude abilities alone are associated with math abilities (Holloway and Ansari, 2009; Vanbinst et al., 2015a). Children in Profile 3 were only poorer on transcoding but not SDA accuracy. It is possible that some children are able to deploy compensatory strategies (e.g., finger-counting) to solve SDA problems and rely less on direct retrieval of arithmetic facts (Geary and Hoard, 2005).

At Time 2, only children with good non-symbolic–symbolic magnitude abilities (i.e., Profile 2) possessed better SDA problem-solving and transcoding. These findings suggest that non-symbolic magnitude abilities are important for symbolic magnitude abilities, and ipso facto, math abilities (i.e., good non-symbolic–symbolic magnitude abilities in Profile 2 at both time points).

Implications and Directions for Future Research

The different change patterns of non-symbolic–symbolic magnitude abilities appear to represent different developmental change pathways. Current findings illustrate a multivariate framework in which different magnitude representation developmental pathways are underpinned by different cognitive factors (VSWM and naming numbers ability) that contribute to differences in math development. Insofar as different magnitude representation change pathways exist, they may reflect typical and atypical models of math development. Nevertheless, since the current study only investigated magnitude abilities over 1 year, caution should be exercised in extrapolating beyond this time period. We are unable to specify what happens to profile membership as children age: it is possible that the profiles will converge on a single magnitude ability competency (i.e., equifinality). It is also possible that differences in profile trajectories will remain separate or diverge further over time. Indeed, an explicit characterization of profile membership over time might help distinguish between delays, differences, deficits in math development. We suggest two of the change pathways (e.g., children transitioning from Profiles 4 to 3 and/or remaining in Profile 3) may well represent an early risk marker for math difficulties (delay and/or deficit). These issues are matter for future research, however. Nonetheless, it should be noted that VSWM and naming numbers ability are important correlates of a typical (and optimal) magnitude representation developmental pathway, and ipso facto, good math abilities.

Our research highlights the value of using LTA for examining data in which more than one developmental trajectories are hypothesized (Lanza and Cooper, 2016). Longitudinal analyses tend to focus on analytical techniques (e.g., correlations, regressions and structural equation models) where the same over time estimates are applied to samples (e.g., Libertus et al., 2013; Vanbinst et al., 2015b). While such analytical models are useful, they may be limited when different developmental trajectories are embedded in a data distribution. In the current study, being able to model different non-symbolic–symbolic magnitude ability change patterns in a single model, along with the associated cognitive factors/math abilities allowed us to represent a more comprehensive approach to modeling development (von Eye and Bergman, 2003).

We note, however, that our magnitude judgment tasks included stimuli from the so-called subitizing ($n \leq 4$) and counting ranges that are thought to depend on different enumeration mechanisms (Reeve et al., 2012). It is possible that including items from the subitizing and counting ranges,

either separately or in combination, may affect judgment responses. However, since similar response patterns occurred for comparison stimuli from the subitizing and counting ranges in both judgment tasks, we suggest the indices are arguably assessing a common underlying construct. Nonetheless, it is not always evident in magnitude judgment tasks whether performance reflects stimulus properties, task demands, or the construct under investigation (see Karolis et al., 2011 for a discussion).

CONCLUSION

The current study replicated and extended Chew et al.'s (2016) findings. Indeed, the interpretive importance of replicating findings in latent class analysis research has recently been strongly emphasized (Hickendorff et al., 2018). The findings showed that identifiable differences in the profiles of relationships between non-symbolic and symbolic magnitude abilities could be extracted from a general distribution of these abilities. It also showed different change/stability pathways in these profiles over 1 year and that these were differentially associated with children's math abilities. And, it also showed that VSWM and naming number abilities were differentially associated with the non-symbolic–symbolic magnitude ability profiles. While the present findings highlight the importance of paying attention to the developmental significance of different patterns of abilities over time which potentially represent typical and atypical developmental models, they should not be over-interpreted. The current study only examined stability/change in non-symbolic and symbolic magnitude abilities over 1 year. The issue of whether pathways converge later in time (i.e., reach equifinality), and the developmental math/cognitive implications of different pathways across extended time, is unable to be addressed in the present study and could usefully be the subject of future research.

AUTHOR CONTRIBUTIONS

CC collected and analyzed the data. RR, JF, and CC collaborated in writing the paper.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2019.00441/full#supplementary-material>

REFERENCES

- Bailey, D. H., Littlefield, A., and Geary, D. C. (2012). The codevelopment of skill at and preference for use of retrieval-based processes for solving addition problems: individual and sex differences from first to sixth grades. *J. Exp. Child Psychol.* 113, 78–92. doi: 10.1016/j.jecp.2012.04.014
- Bakk, Z., Tekle, F. B., and Vermunt, J. K. (2013). Estimating the association between latent class membership and external variables using bias-adjusted three-step approaches. *Sociol. Methodol.* 43, 272–311. doi: 10.1177/0081175012470644
- Bartelet, D., Vaessen, A., Blomert, L., and Ansari, D. (2014). What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *J. Exp. Child Psychol.* 117, 12–28. doi: 10.1016/j.jecp.2013.08.010
- Bergman, L. R., Magnusson, D., and El-Khoury, B. M. (2003). *Studying Individual Development in an Interindividual Context: A Person-Oriented Approach*.

- Paths through Life*, 1st Edn, Vol. 4. London: Psychology Press. doi: 10.1017/CBO9781107415324.004
- Berteletti, I., Lucangeli, D., Piazza, M., Dehaene, S., and Zorzi, M. (2010). Numerical estimation in preschoolers. *Dev. Psychol.* 46, 545–551. doi: 10.1037/a0017887
- Bouwmeester, S., and Verkoeijen, P. P. J. L. (2012). Multiple representations in number line estimation: a developmental shift or classes of representations? *Cogn. Instr.* 30, 246–260. doi: 10.1080/07370008.2012.689384
- Bugden, S., and Ansari, D. (2011). Individual differences in children's mathematical competence are related to the intentional but not automatic processing of Arabic numerals. *Cognition* 118, 35–47. doi: 10.1016/j.cognition.2010.09.005
- Chen, Q., and Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychol.* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Chew, C. S., Forte, J. D., and Reeve, R. A. (2016). Cognitive factors affecting children's nonsymbolic and symbolic magnitude judgment abilities: a latent profile analysis. *J. Exp. Child Psychol.* 152, 173–191. doi: 10.1016/j.jecp.2016.07.001
- Clark, S., and Muthén, B. (2009). *Relating Latent Class Analysis Results to Variables not Included in the Analysis*. Available at: <http://www.statmodel.com/download/relatinglca.pdf>
- Collins, L. M., and Lanza, S. T. (2013). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Sciences*. Hoboken, NJ: John Wiley & Sons.
- Cotton, S. M., Kiely, P. M., Crewther, D. P., Thomson, B., Laycock, R., and Crewther, S. G. (2005). A normative and reliability study for the Raven's Coloured Progressive Matrices for primary school aged children from Victoria, Australia. *Pers. Individ. Dif.* 39, 647–659. doi: 10.1016/j.paid.2005.02.015
- de Hevia, M. D. (2016). *Core Mathematical Abilities in Infants: Number and Much More. Progress in Brain Research*, 1st Edn. Amsterdam: Elsevier B.V. doi: 10.1016/bs.pbr.2016.04.014
- de Hevia, M. D., Vallar, G., and Girelli, L. (2008). Visualizing numbers in the mind's eye: the role of visuo-spatial processes in numerical abilities. *Neurosci. Biobehav. Rev.* 32, 1361–1372. doi: 10.1016/j.neubiorev.2008.05.015
- De Smedt, B., and Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *J. Exp. Child Psychol.* 108, 278–292. doi: 10.1016/j.jecp.2010.09.003
- De Smedt, B., Noël, M.-P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends Neurosci. Educ.* 2, 48–55. doi: 10.1016/j.tine.2013.06.001
- De Smedt, B., Verschaffel, L., and Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *J. Exp. Child Psychol.* 103, 469–479. doi: 10.1016/j.jecp.2009.01.010
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N
- Dehaene, S. (2007). "Symbols and quantities in parietal cortex: elements of a mathematical theory of number representation and manipulation," in *Attention & Performance XXII. Sensori-Motor Foundation of Higher Cognition*, eds P. Haggard and Y. Rossetti (Cambridge, MA: Harvard University Press), 527–574. doi: 10.1093/acprof:oso/9780199231447.003.0024
- Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition*. Oxford: Oxford University Press.
- Dehaene, S., and Brannon, E. M. (2011). *Space, Time and Number in the Brain: Searching for the Foundations of Mathematical Thought*. London: Academic Press.
- Dehaene, S., and Cohen, L. (1997). Cerebral pathways for calculation: double dissociation between rote verbal and quantitative knowledge of arithmetic. *Cortex* 33, 219–250. doi: 10.1016/S0010-9452(08)70002-9
- Dehaene, S., Izard, V., and Piazza, M. (2005). *Control Over Non-Numerical Parameters in Numerosity Experiments*. Available at: [Www.Unicon.Org](http://www.unicon.org)
- Dehaene, S., Piazza, M., Pinel, P., and Cohen, L. (2003). Three parietal circuits for number processing. *Cogn. Neuropsychol.* 20, 487–506. doi: 10.1080/02643290244000239
- Desoete, A., Ceulemans, A., De Weerd, F., and Pieters, S. (2012). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study. *Br. J. Educ. Psychol.* 82, 64–81. doi: 10.1348/2044-8279.002002
- Dowker, A. (2008). Individual differences in numerical abilities in preschoolers. *Dev. Sci.* 11, 650–654. doi: 10.1111/j.1467-7687.2008.00713.x
- Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123C, 53–72. doi: 10.1016/j.jecp.2014.01.013
- Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002
- Feigenson, L., Libertus, M. E., and Halberda, J. (2013). Links between the intuitive sense of number and formal mathematics ability. *Child Dev. Perspect.* 7, 74–79. doi: 10.1111/cdep.12019
- Friso-Van Den Bos, I., Van Der Ven, S. H. G., Kroesbergen, E. H., and Van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: a meta-analysis. *Educ. Res. Rev.* 10, 29–44. doi: 10.1016/j.edurev.2013.05.003
- Geary, D. C. (2000). From infancy to adulthood: the development of numerical abilities. *Eur. Child Adolesc. Psychiatry* 9(Suppl. 2), 11–16. doi: 10.1007/s007870070004
- Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Dev. Psychol.* 47, 1539–1552. doi: 10.1037/a0025510
- Geary, D. C., and Hoard, M. K. (2005). "Learning disabilities in arithmetic and mathematics: theoretical and empirical perspectives," in *Handbook of Mathematical Cognition*, ed. J. I. D. Campbell (London: Psychology Press), 253–266.
- Gebuis, T., Cohen Kadosh, R., De Haan, E., and Henik, A. (2009). Automatic quantity processing in 5-year olds and adults. *Cogn. Process.* 10, 133–142. doi: 10.1007/s10339-008-0219-x
- Gray, S. A., and Reeve, R. A. (2014). Preschoolers' dot enumeration abilities are markers of their arithmetic competence. *PLoS One* 9:e94428. doi: 10.1371/journal.pone.0094428
- Gray, S. A., and Reeve, R. A. (2016). Number-specific and general cognitive markers of preschoolers' math ability profiles. *J. Exp. Child Psychol.* 147, 1–21. doi: 10.1016/j.jecp.2016.02.004
- Halberda, J., Mazocco, M. M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* 455, 665–668. doi: 10.1038/nature07246
- Hickendorff, M., Edelsbrunner, P. A., McMullen, J., Schneider, M., and Trezise, K. (2018). Informative tools for characterizing individual differences in learning: latent class, latent profile, and latent transition analysis. *Learn. Individ. Dif.* 66, 4–15. doi: 10.1016/j.lindif.2017.11.001
- Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children's mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001
- Izard, V., Sann, C., Spelke, E. S., and Streri, A. (2009). Newborn infants perceive abstract numbers. *Proc. Natl. Acad. Sci. U.S.A.* 106, 10382–10385. doi: 10.1073/pnas.0812142106
- Jordan, N. C., Hansen, N., Fuchs, L. S., Siegler, R. S., Gersten, R., and Micklos, D. (2013). Developmental predictors of fraction concepts and procedures. *J. Exp. Child Psychol.* 116, 45–58. doi: 10.1016/j.jecp.2013.02.001
- Karolis, V., Iuculano, T., and Butterworth, B. (2011). Mapping numerical magnitudes along the right lines: differentiating between scale and bias. *J. Exp. Psychol. Gen.* 140, 693–706. doi: 10.1037/a0024255
- Kessels, R. P., van Zandvoort, M. J., Postma, A., Kappelle, L. J., and de Haan, E. H. (2000). The Corsi block-tapping task: standardization and normative data. *Appl. Neuropsychol.* 7, 252–258. doi: 10.1207/S15324826AN0704_8
- Kolkman, M. E., Kroesbergen, E. H., and Leseman, P. P. M. (2013). Early numerical development and the role of non-symbolic and symbolic skills. *Learn. Instr.* 25, 95–103. doi: 10.1016/j.learninstruc.2012.12.001
- Kyttälä, M., and Lehto, J. E. (2008). Some factors underlying mathematical performance: the role of visuospatial working memory and non-verbal intelligence. *Eur. J. Psychol. Educ.* 23, 77–94. doi: 10.1007/BF03173141
- Lanza, S. T., and Cooper, B. R. (2016). Latent class analysis for developmental research. *Child Dev. Perspect.* 10, 59–64. doi: 10.1111/cdep.12163

- Laursen, B. P., and Hoff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill Palmer Q.* 52, 377–389. doi: 10.1353/mpq.2006.0029
- Libertus, M. E., Feigenson, L., and Halberda, J. (2013). Is approximate number precision a stable predictor of math ability? *Learn. Individ. Dif.* 25, 126–133. doi: 10.1016/j.lindif.2013.02.001
- Luwel, K., Foustana, A., Onghena, P., and Verschaffel, L. (2013). The role of verbal and performance intelligence in children's strategy selection and execution. *Learn. Individ. Dif.* 24, 134–138. doi: 10.1016/j.lindif.2013.01.010
- Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., and Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: the case of the numerical distance effect. *Acta Psychol.* 134, 154–161. doi: 10.1016/j.actpsy.2010.01.006
- Mammarella, I. C., Lucangeli, D., and Cornoldi, C. (2010). Spatial working memory and arithmetic deficits in children with nonverbal learning difficulties. *J. Learn. Disabil.* 43, 455–468. doi: 10.1177/0022219409355482
- Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x
- Moyer, R. S., and Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature* 215, 1519–1520. doi: 10.1038/2151519a0
- Muthen, B., and Muthen, L. K. (2000). Integrating person-centered and variable-centered analyses: growth mixture modeling with latent trajectory classes. *Alcohol. Clin. Exp. Res.* 24, 882–891. doi: 10.1111/j.1530-0277.2000.tb02070.x
- OECD (2012). *Education at a Glance, Education at a Glance 2012: OECD Indicators*. Paris: OECD Publishing. doi: 10.1787/eag-2012-en
- Paul, J. M., Gray, S. A., Butterworth, B. L., and Reeve, R. A. (2018). Reading and math tests differentially predict number transcoding and number fact speed longitudinally: a random intercept cross-lagged panel approach. *J. Educ. Psychol.* 111, 299–313. doi: 10.1037/edu0000287
- Paul, J. M., and Reeve, R. A. (2016). Relationship between single digit addition strategies and working memory reflects general reasoning sophistication. *Learn. Instr.* 42, 113–122. doi: 10.1016/j.learninstruc.2016.01.011
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends Cogn. Sci.* 14, 542–551. doi: 10.1016/j.tics.2010.09.008
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012
- Piazza, M., and Izard, V. (2009). How humans count: numerosity and the parietal cortex. *Neuroscientist* 15, 261–273. doi: 10.1177/1073858409333073
- Price, G. R., Palmer, D., Battista, C., and Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychol.* 140, 50–57. doi: 10.1016/j.actpsy.2012.02.008
- Ratcliff, R., Thompson, C. A., and Mckoon, G. (2015). Modeling individual differences in response time and accuracy in numeracy. *Cognition* 137, 115–136. doi: 10.1016/j.cognition.2014.12.004
- Raven, J., Court, J. H., and Raven, J. C. (1995). *Coloured Progressive Matrices*. Oxford: Oxford Psychologists Press.
- Reeve, R., Gray, S., Butterworth, B., and Paul, J. (2018). Variability in single digit addition problem-solving speed over time identifies typical, delay and deficit math pathways. *Front. Psychol.* 9:1498. doi: 10.3389/fpsyg.2018.01498
- Reeve, R., Reynolds, F., Humberstone, J., and Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *J. Exp. Psychol. Gen.* 141, 649–666. doi: 10.1037/a0027520
- Rousselle, L., and Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs non-symbolic number magnitude processing. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005
- Sasanguie, D., De Smedt, B., Defever, E., and Reynvoet, B. (2012a). Association between basic numerical abilities and mathematics achievement. *Br. J. Dev. Psychol.* 30, 344–357. doi: 10.1111/j.2044-835X.2011.02048.x
- Sasanguie, D., Van den Bussche, E., and Reynvoet, B. (2012b). Predictors for mathematics achievement? Evidence from a longitudinal study. *Mind Brain Educ.* 6, 119–128. doi: 10.1111/j.1751-228X.2012.01147.x
- Sasanguie, D., Defever, E., Maertens, B., and Reynvoet, B. (2014). The approximate number system is not predictive for symbolic number processing in kindergarteners. *Q. J. Exp. Psychol.* 67, 271–280. doi: 10.1080/17470218.2013.803581
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: what underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418–431. doi: 10.1016/j.jecp.2012.10.012
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372
- Siegler, R. S. (1987). The perils of averaging data over strategies: an example from children's addition. *J. Exp. Psychol. Gen.* 116, 250–264. doi: 10.1037/0096-3445.116.3.250
- Siegler, R. S. (2016). Continuity and change in the field of cognitive development and in the perspectives of one cognitive developmentalist. *Child Dev. Perspect.* 10, 128–133. doi: 10.1111/cdep.12173
- Träff, U. (2013). The contribution of general cognitive abilities and number abilities to different aspects of mathematics in children. *J. Exp. Child Psychol.* 116, 139–156. doi: 10.1016/j.jecp.2013.04.007
- Treize, K., and Reeve, R. A. (2014). Cognition-emotion interactions: patterns of change and implications for math problem solving. *Front. Psychol.* 5:840. doi: 10.3389/fpsyg.2014.00840
- Vanbinst, K., Ceulemans, E., Ghesquière, P., and De Smedt, B. (2015a). Profiles of children's arithmetic fact development: a model-based clustering approach. *J. Exp. Child Psychol.* 133, 29–46. doi: 10.1016/j.jecp.2015.01.003
- Vanbinst, K., Ghesquière, P., and De Smedt, B. (2015b). Does numerical processing uniquely predict first graders' future development of single-digit arithmetic? *Learn. Individ. Differ.* 37, 153–160. doi: 10.1016/j.lindif.2014.12.004
- Vermunt, J. K., and Magidson, J. (2013a). *Latent Gold 5.0 Upgrade Manual 1*. Belmont, MA: Statistical Innovations Inc.
- Vermunt, J. K., and Magidson, J. (2013b). *Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations Inc.
- Vermunt, J. K., and Magidson, J. (2015). *Upgrade Manual for Latent GOLD 5.1*. Belmont, MA: Statistical Innovations.
- von Eye, A., and Bergman, L. R. (2003). Research strategies in developmental psychopathology: dimensional identity and the person-oriented approach. *Dev. Psychopathol.* 15, 553–580. doi: 10.1017/S0954579403000294
- Xenidou-Dervou, I., Molenaar, D., Ansari, D., van der Schoot, M., and van Lieshout, E. C. D. M. (2016). Nonsymbolic and symbolic magnitude comparison skills as longitudinal predictors of mathematical achievement. *Learn. Instr.* 50, 1–13. doi: 10.1016/j.learninstruc.2016.11.001
- Zorzi, M., Priftis, K., and Umiltà, C. (2002). Brain damage: neglect disrupts the mental number line. *Nature* 417, 138–139. doi: 10.1038/417138a

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Chew, Forte and Reeve. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Differences in Counting Skills Between Chinese and German Children Are Accompanied by Differences in Processing of Approximate Numerical Magnitude Information

Jan Lonnemann^{1,2,3*}, Su Li^{4,5}, Pei Zhao^{4,5,6}, Janosch Linkersdörfer^{2,3}, Sven Lindberg^{3,7}, Marcus Hasselhorn^{2,3,8} and Song Yan⁹

OPEN ACCESS

Edited by:

Xinlin Zhou,
Beijing Normal University, China

Reviewed by:

Frank Domahs,
University of Marburg, Germany
Jo Van Herwegen,
Kingston University, United Kingdom
Jennifer B. Wagner,
College of Staten Island,
United States

*Correspondence:

Jan Lonnemann
lonnemann@uni-potsdam.de

Specialty section:

This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

Received: 15 May 2018

Accepted: 10 December 2018

Published: 08 January 2019

Citation:

Lonnemann J, Li S, Zhao P,
Linkersdörfer J, Lindberg S,
Hasselhorn M and Yan S (2019)
Differences in Counting Skills
Between Chinese and German
Children Are Accompanied by
Differences in Processing of
Approximate Numerical
Magnitude Information.
Front. Psychol. 9:2656.
doi: 10.3389/fpsyg.2018.02656

¹Empirical Childhood Research, University of Potsdam, Potsdam, Germany, ²Department of Education and Human Development, Leibniz Institute for Research and Information in Education (DIPF), Frankfurt am Main, Germany, ³Center for Individual Development and Adaptive Education of Children at Risk (IDeA), Frankfurt am Main, Germany, ⁴Institute for Psychology, Chinese Academy of Sciences, Beijing, China, ⁵Department of Psychology, University of Chinese Academy of Sciences (UCAS), Beijing, China, ⁶Faculty of Education, Beijing City University, Beijing, China, ⁷Faculty of Arts and Humanities, University of Paderborn, Paderborn, Germany, ⁸Department of Educational Psychology, Goethe-Universität Frankfurt am Main, Frankfurt am Main, Germany, ⁹Department of Psychology and Methods, Jacobs University Bremen, Bremen, Germany

Human beings are supposed to possess an approximate number system (ANS) dedicated to extracting and representing approximate numerical magnitude information as well as an object tracking system (OTS) for the rapid and accurate enumeration of small sets. It is assumed that the OTS and the ANS independently contribute to the acquisition of more elaborate numerical concepts. Chinese children have been shown to exhibit more elaborate numerical concepts than their non-Chinese peers, but it is still an open question whether similar cross-national differences exist with regard to the underlying systems, namely the ANS and the OTS. In the present study, we investigated this question by comparing Chinese and German preschool children with regard to their performance in a non-symbolic numerical magnitude comparison task (assessing the ANS) and in an enumeration task (assessing the OTS). In addition, we compared children's counting skills. To ensure that possible between-group differences could not be explained by differences in more general performance factors, we also assessed children's reasoning ability and processing speed. Chinese children showed a better counting performance and a more accurate performance in the non-symbolic numerical magnitude comparison task. These differences in performance could not be ascribed to differences in reasoning abilities and processing speed. In contrast, Chinese and German children did not differ significantly in the enumeration of small sets. The superior counting performance of Chinese children was thus found to be reflected in the ANS but not in the OTS.

Keywords: approximate number system, subitizing, counting, cross-national comparison, preschool

INTRODUCTION

Human beings are assumed to possess an evolutionarily ancient, innate system dedicated to extracting and representing approximate numerical magnitude information. This so-called approximate number system (ANS; see Piazza, 2010, for an overview) enables us to discriminate between sets of different quantities and is proposed to serve as the foundation for the acquisition of more elaborate numerical concepts (e.g., Feigenson et al., 2004). We are faster and more accurate in comparing two visually presented dot arrays with respect to their quantity the more their ratio deviates from one (e.g., van Oeffelen and Vos, 1982). The ability to discriminate between sets of different numerical quantities seems to already exist in preverbal infants (e.g., Izard et al., 2009) and undergoes a progressive refinement throughout development (Piazza, 2010; Halberda et al., 2012). Besides this developmental variation, individuals of the same age show inter-individual differences in their ability to discriminate between sets of numerical quantities. Recent meta-analyses demonstrated that these differences are linked to symbolic math performance (Chen and Li, 2014; Fazio et al., 2014; Schneider et al., 2017). According to Chen and Li (2014), this association remains significant even when considering potential moderators like general cognitive abilities, and it is comparable in strength in children and adults. On the other hand, Fazio et al. (2014) reported higher correlations for children than for adults and Schneider et al. (2017) also detected a similar but small moderating effect of age.

In addition to the ANS, a so-called object tracking system (OTS; see e.g., Piazza, 2010, for an overview) has been proposed. The OTS is assumed to enable “subitizing,” i.e., the rapid and accurate judgment of the number of small sets “at a glance,” without counting. Indeed, children can determine the number of objects in small sets of three or four items with high speed and high accuracy (Pylyshyn, 2001; Revkin et al., 2008). Similar to the ANS, the OTS undergoes a refinement throughout development and shows inter-individual differences (e.g., Reeve et al., 2012). The OTS is assumed to independently contribute to the acquisition of more elaborate numerical concepts (Feigenson et al., 2004). This is supported by studies showing an association between the ability to rapidly and accurately enumerate small sets with concurrent and future math achievement (e.g., Reeve et al., 2012; Gray and Reeve, 2014; Major et al., 2017). Dot enumeration tasks are typically used to assess the OTS. In these tasks, different sets of dots are presented (e.g., 1–9 dots) and the participants are asked to verbally state as quickly and as correctly as possible the respective number of dots. Based on a typical response pattern with a relatively flat slope for small sets of dots (1–3/4) and a steeper slope for larger sets of dots (4/5–9), it is assumed that at least two distinct systems are involved: a subitizing system (OTS) and a counting system (see e.g., Major et al., 2017). According to Piazza (2010), the number of objects in sets with more than three or four items can indeed only be assessed using exact counting or approximate estimation.

Cross-national assessments of mathematical achievement have repeatedly demonstrated that Chinese children outperform

their non-Chinese peers at various ages (e.g., Wang and Lin, 2009, 2013; Mullis et al., 2012; OECD, 2013). This superior Chinese performance has been attributed to different factors including number naming systems, cultural beliefs and values, parental involvement, as well as educational systems and practices (Ng and Rao, 2010). Cross-national differences seem to emerge even before children enter elementary school. A study by Miller et al. (1995), for example, revealed that 4-year-old Chinese children can count much higher than their American peers. Moreover, Aunio et al. (2008) compared Chinese, English, and Finnish preschool children’s performance in the Early Numeracy Test (ENT; Van Luit et al., 1994). According to the authors, the ENT assesses children’s use and understanding of numbers (so-called counting skills) as well as children’s understanding of quantities and relations (so-called relational skills). Counting skills were assessed by probing children’s knowledge of cardinal and ordinal numbers up to 20 (e.g., “Count on from 9 to 15”). Relational skills were assessed by asking children to compare two non-equivalent cardinal or ordinal situations from given pictures (e.g., “Here you see Indians. Point out the Indian who has less feathers than this Indian with bow and arrow”). Chinese children showed better counting skills and better relational skills than their non-Chinese peers (Aunio et al., 2008). In a related study with 4- to 7-year-old participants, Chinese children showed better counting skills than Finnish children irrespective of age, whereas only older Chinese children outperformed their Finnish counterparts in relational skills (Aunio et al., 2006). In sum, there exists ample evidence that Chinese children have more elaborated numerical concepts than their non-Chinese peers. Whether similar cross-national differences exist with regard to the ANS and the OTS, however, remains an open question.

To the best of our knowledge, there is only one study investigating differences in the ANS between Chinese and non-Chinese preschool children. Rodic et al. (2014) compared 5- to 7-year-old children from China, Kyrgyzstan, Russia, and the UK. They assessed simple arithmetic skills, the ANS, and other skills assumed to be related to the development of arithmetic skills, i.e., number naming, symbolic numerical magnitude comparison, and dot enumeration. The dot enumeration task evaluated children’s ability to map a number of dots to Arabic numerals and therefore did not directly assess the OTS. While the Chinese children significantly outperformed all other groups in the arithmetic tasks, this result was not (exactly) mirrored in the non-symbolic numerical magnitude comparison task (assessing the ANS). While Chinese children showed better non-symbolic numerical magnitude comparison performance than UK, Dungan, and Kyrgyz children, they did not significantly outperform Russian children. According to Rodic et al. (2014), the observed small advantage of Chinese and Russian children in the non-symbolic numerical magnitude comparison task supports the view that the link between the ANS and mathematical skills is relatively weak and potentially reversed (mathematical skills affecting the ANS). Meta-analytic findings by Chen and Li (2014) provide evidence for both directions of influence: non-symbolic numerical magnitude

processing skills predict later math performance ($r = 0.24$, based on six longitudinal samples), but they can also be predicted by earlier math performance ($r = 0.17$, based on five longitudinal samples).

While the abovementioned findings show that Chinese children have better arithmetic skills and more elaborate numerical concepts than their non-Chinese peers, they do not deliver any clear evidence as to whether the proposed underlying systems, namely the ANS and the OTS, are more elaborate in Chinese children than in their non-Chinese peers. In the present study, we investigated this question by comparing Chinese and German preschool children with regard to their performance in a non-symbolic numerical magnitude comparison task (assessing the ANS) and in an enumeration task (assessing the OTS). In addition, we compared children's counting skills. To assure that possible between-group differences could not be ascribed to differences in more general performance factors, we also assessed reasoning abilities and processing speed. We did not assume that Chinese and German children differ in their ANS/OTS independently of their learning experience. Based on the assumption that mathematical learning affects children's ANS (see Rodic et al., 2014), we hypothesized that Chinese children not only have better counting skills than their German peers but also have better non-symbolic numerical magnitude processing skills. With regard to the OTS, we did not expect any difference between Chinese and German children, as we were not aware of evidence for an influence of mathematical learning experiences on the OTS.

MATERIALS AND METHODS

Participants

The German sample consisted of 37 children (20 females, mean age 60 months, range 49–74 months) recruited from different kindergartens in the region of Frankfurt am Main. The Chinese sample consisted of 37 children (18 females, mean age 59 months, range 48–70 months) recruited from different kindergartens in the region of Beijing. Written and informed consent was obtained from the parents of all participating children. Children additionally provided verbal assent to participate in the study and were compensated for participation (e.g., by receiving a pencil). Our study was not approved by an ethics committee. This is due to the fact that data acquisition for our study started at a time when it was not common practice to apply for an ethics committee approval for psychological studies involving only cognitive measures like ours.

Procedure

All participants were tested individually and performed the tasks in the following order: non-symbolic numerical magnitude comparison, enumeration, processing speed, counting, and reasoning. Computerized tasks (non-symbolic numerical magnitude comparison, enumeration, and processing speed) were programmed and controlled using Presentation® software (Neurobehavioral Systems, Inc.)

Non-symbolic Numerical Magnitude Comparison Task

Sets of black dots were presented in two white squares on the left- and the right-hand sides of the screen. On each trial, one of the white squares contained 32 dots (reference numerosities) and the other one 14, 20, 26, 38, 44, or 50 dots (deviants). This resulted in six different comparison pairs. Each of the six comparison pairs appeared eight times, four times with the reference numerosity on the left and four times on the right-hand side. Every single comparison pair had a unique configuration of dots. The dot sets were created using a Matlab script by Gebuis and Reynvoet (2011) which varied different visual properties of the stimuli [i.e., area extended (convex hull), total surface (the aggregate surface of all dots in one array), density (area extended/total surface), item size (average diameter of the dots presented in one array), and total circumference (circumference of all dots in one array, taken together)] so that no single visual cue was informative about numerical magnitude across all trials. Each of the five different visual cue conditions involved trials in which the respective visual cue was congruent or incongruent with the numerical dimension. Children were asked to indicate, without using counting strategies, the side of the larger numerical magnitude by pressing the left CTRL-button of the computer keyboard with their left index finger when it was larger on the left-hand side and by pressing the right CTRL-button using their right index finger when it was larger on the right-hand side. Reaction times (RT) and errors (ER) were recorded, and the instruction stressed both speed and accuracy. The order of trials was pseudo-randomized to avoid consecutive identical comparison pairs. The experiment started with six warm-up trials (stimuli: 50 vs. 32, 32 vs. 14, 26 vs. 32, 38 vs. 32, 32 vs. 44, 20 vs. 32; no feedback, data not recorded), followed by 48 experimental trials (6 comparison pairs \times 8 repetitions). The experimenter pressed a button to start a trial, whereupon a black screen was presented for 1,000 ms. After the black screen had vanished, the target appeared until a response was given, but only up to a maximum duration of 6,000 ms. If no response was given, a trial was classified as erroneous. No feedback was given regarding the correctness of responses. Mean RT and mean ER were used as individual markers of the ANS (see, e.g., Inglis and Gilmore, 2014, for a discussion on different indices of the ANS). Correct responses were used for computing mean RT. Response times below 200 ms were excluded from further analysis. This trimming resulted in 0.06% of response exclusions for Chinese participants and in 0.28% of response exclusions for German participants.

Enumeration

Sets of dots were presented in a white square in the center of the screen. On each trial, the white square contained 1, 2, 3, 4, 5, 6, 7, 8, or 9 dots. Each number of dots appeared two times and every single stimulus had a unique configuration of dots. Children were asked to verbally state as quickly and as correctly as possible the respective number of dots. To assess RT, the examiner pressed a button on an external device as soon as the child began to verbalize the answer. Then, a black screen appeared, while the examiner

recorded the answer given by the child. Afterward, a new stimulus was presented. Targets appeared until the child gave an answer. No feedback was provided regarding the correctness of responses. The experiment started with four warm-up trials (stimuli: 4, 2, 8, 5; no feedback, data not recorded), followed by 18 experimental trials in total. The order of trials was pseudo-randomized so that the number of dots was not identical on consecutive trials. Mean RT and ER as well as RT slopes for sets of dots in the subitizing range were used as individual markers of the OTS. Correct responses were used for computing mean RT. ER in the subitizing range can be assumed to be very low, but from our point of view, it is still important to consider ER in the subitizing range, since it cannot be excluded from the outset that there are no group differences in this respect. In addition, mean ER as well as ER slopes for the enumeration of sets of dots beyond the subitizing range were analyzed.

Counting

Children were asked to recite the number word sequence from 1 to 30. The last number that was counted correctly was used to estimate children's counting skills.

Reasoning

Raven's Colored Progressive Matrices (CPM; Bultheller and Häcker, 2002) were used to assess inductive reasoning. The CPM is an untimed power test consisting of 36 colored diagrammatic puzzles, each with a missing part which has to be identified from a choice of six. Total scores ranging from 0 to 36 are reported for each child.

Processing Speed

A visual detection task was used to assess individual processing speed. Children were instructed to press the space bar of the computer's keyboard as soon as possible whenever an "X" appeared in the center of the screen. The target appeared until a response was given, but only up to a maximum duration of 3,000 ms. The task comprised 10 experimental trials with varying inter-trial intervals (2,000, 3,500, 5,000, 6,500, or 8,000 ms). Correct responses were used for computing mean RT. If no response was given, a trial was classified as erroneous. Mean ER in the visual detection task was low (Chinese children: 0.0%; German children: 0.5%) and not further analyzed.

Analyses

To assess the effect of ratio between the two to-be-compared numerical magnitudes in the non-symbolic numerical magnitude comparison task, we collapsed trials with deviants smaller than the reference (14, 20, 26) and trials with deviants larger than the reference (38, 44, 50) into three levels of ratio [14/50 vs. 32 (ratios = 0.4375/1.5625), 20/44 vs. 32 (ratios = 0.625/1.375), and 26/38 vs. 32 (ratios = 0.8125/1.1875)] and used polynomial linear trend analyses for collapsed ratios separately for ER and RT. Moreover, we used two-sample *t*-tests to assess differences between Chinese and German children with regard to age, reasoning, processing speed, counting skills, mean RT/ER in the non-symbolic numerical magnitude comparison task, as well as mean RT/ER and RT slopes for the enumeration of sets of dots in the subitizing range. In subsequent analyses, we used two-sample *t*-tests to compare Chinese and German children with regard to mean ER and ER slopes for the enumeration

of sets of dots beyond the subitizing range. The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

RESULTS

Demonstrating the signature of the ANS, ER in the non-symbolic numerical magnitude comparison task decreased the more the ratio between the two to-be-compared numerosities deviated from one [Chinese children: 26/38 vs. 32: ER = 45%, 20/44 vs. 32: ER = 31%, 14/50 vs. 32: ER = 23%; $F(1, 36) = 72.85$, $p < 0.001$, $\eta_p^2 = 0.67$; German children: 26/38 vs. 32: ER = 47%, 20/44 vs. 32: ER = 36%, 14/50 vs. 32: ER = 29%; $F(1, 36) = 75.75$, $p < 0.001$, $\eta_p^2 = 0.68$]. On the basis of RT in the non-symbolic numerical magnitude task, a significant linear trend was found for German children [26/38 vs. 32: RT = 1,394 ms, 20/44 vs. 32: RT = 1,641 ms, 14/50 vs. 32: RT = 1,503 ms; $F(1, 36) = 10.48$, $p < 0.01$, $\eta_p^2 = 0.23$] but not for Chinese children [26/38 vs. 32: RT = 1,327 ms, 20/44 vs. 32: RT = 1,422 ms, 14/50 vs. 32: RT = 1,371 ms; $F(1, 36) = 0.79$, $p = 0.379$, $\eta_p^2 = 0.02$]. German children unexpectedly showed fastest RT when the ratio between the two to-be-compared numerosities was least different from one (26 or 38 vs. 32).¹ There was, however, no indication of a speed-accuracy trade-off in German children ($r = 0.26$, $p = 0.123$).

In the enumeration task, some children did not respond correctly in all trials of a specific condition and thus RT for correct responses could not be determined for all participants in each condition (see **Figure 1**). When considering RT for correct responses as well as ER in the enumeration task, both Chinese and German children showed a typical response pattern (see **Figure 1**), with a relatively flat slope for small sets of dots (1–3) and a steeper slope for larger sets of dots (4–9). We interpreted these results as an indication for a subitizing range of 1–3 in both groups of children. Within the subitizing range, it was possible to determine RT for correct enumerations in each of the different conditions. Mean RT and ER

¹Similar results were found when trials with deviants smaller than the reference (26, 20, 14) and trials with deviants larger than the reference (38, 44, 50) were analyzed separately: ER decreased the more the ratio between the two to-be-compared numerosities deviated from one for deviants smaller than the reference [Chinese children: 26 vs. 32: ER = 46%, 20 vs. 32: ER = 22%, 14 vs. 32: ER = 14%; $F(1, 36) = 76.30$, $p < 0.001$, $\eta_p^2 = 0.68$; German children: 26 vs. 32: ER = 48%, 20 vs. 32: ER = 27%, 14 vs. 32: ER = 19%; $F(1, 36) = 92.13$, $p < 0.001$, $\eta_p^2 = 0.72$] as well as for deviants larger than the reference [Chinese children: 38 vs. 32: ER = 44%, 44 vs. 32: ER = 40%, 50 vs. 32: ER = 31%; $F(1, 36) = 16.82$, $p < 0.001$, $\eta_p^2 = 0.32$; German children: 38 vs. 32: ER = 47%, 44 vs. 32: ER = 46%, 50 vs. 32: ER = 40%; $F(1, 36) = 5.97$, $p < 0.05$, $\eta_p^2 = 0.14$]. On the basis of RT, a significant linear trend was found for German children in the case of deviants smaller than the reference [26 vs. 32: RT = 1,391 ms, 20 vs. 32: RT = 1,605 ms, 14 vs. 32: RT = 1,577 ms; $F(1, 36) = 15.72$, $p < 0.001$, $\eta_p^2 = 0.23$]. German children showed fastest RT when the ratio between the two to-be-compared numerosities was least different from one (26 vs. 32). No significant linear trends were found in the other conditions [Chinese children, deviants smaller than the reference: 26 vs. 32: RT = 1,402 ms, 20 vs. 32: RT = 1,402 ms, 14 vs. 32: RT = 1,422 ms; $F(1, 36) = 0.96$, $p = 0.759$, $\eta_p^2 = 0.003$; Chinese children, deviants larger than the reference: 38 vs. 32: RT = 1,253 ms, 44 vs. 32: RT = 1,442 ms, 50 vs. 32: RT = 1,320 ms; $F(1, 36) = 1.63$, $p = 0.210$, $\eta_p^2 = 0.04$; German children, deviants larger than the reference: 38 vs. 32: RT = 1,397 ms, 44 vs. 32: RT = 1,678 ms, 50 vs. 32: RT = 1,428 ms; $F(1, 36) = 0.43$, $p = 0.518$, $\eta_p^2 = 0.01$].

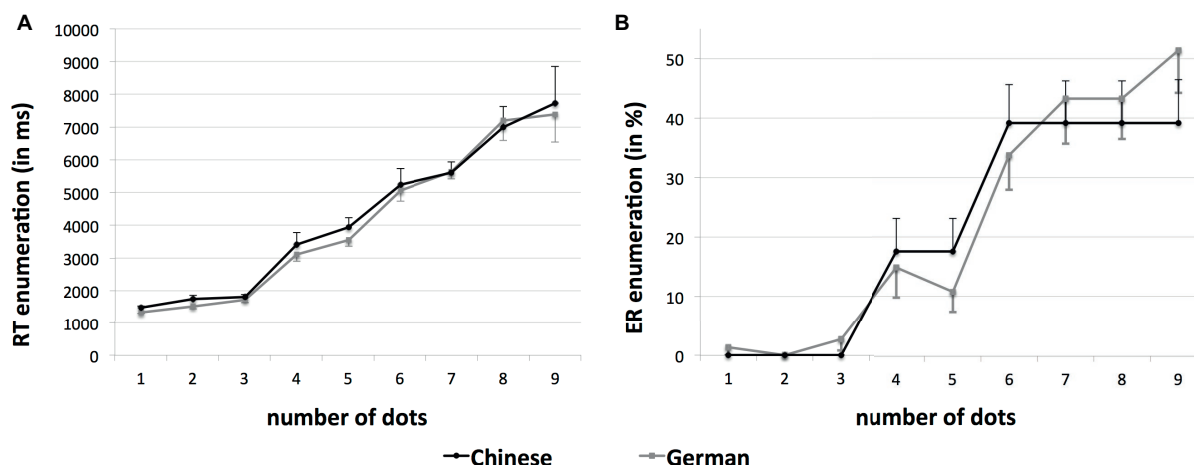


FIGURE 1 | Reaction times (RT) for correct responses and error rates (ER) in the enumeration task. **(A)** RT (in ms) separately for Chinese and German children as a function of the number of dots. The sample size varies depending on the condition (number of dots), because some children did not respond correctly in all trials of a specific condition and thus RT for correct responses could not be determined. The sample size of Chinese and German children of the different conditions is as follows: number of dots = 1, 37 Chinese and 37 German children; number of dots = 2, 37 Chinese and 37 German children; number of dots = 3, 37 Chinese and 37 German children; number of dots = 4, 33 Chinese and 34 German children; number of dots = 5, 33 Chinese and 37 German children; number of dots = 6, 29 Chinese and 32 German children; number of dots = 7, 27 Chinese and 24 German children; number of dots = 8, 27 Chinese and 27 German children; and number of dots = 9, 26 Chinese and 23 German children. **(B)** ER (in %) separately for Chinese and German children as a function of the number of dots. Error bars depict one standard error of the mean.

as well as the best-fitting regression lines for each child's RT were calculated for this range. Beyond the subitizing range (4–9), it was not possible to determine RT for correct enumerations in each of the different conditions. Accordingly, ER (in %) as well as ER slopes were calculated for this range.

While Chinese and German children did not differ significantly with regard to age [$t(72) = 0.13$, $p = 0.897$, $r = 0.02$], reasoning abilities [$t(64) = -1.39$, $p = 0.168$, $r = -0.16$], or processing speed [$t(72) = -0.87$, $p = 0.388$, $r = -0.10$], Chinese children were able to count significantly higher [$t(72) = -3.16$, $p = 0.002$, $r = -0.34$]. This superior counting performance of Chinese children was accompanied by a higher accuracy in the non-symbolic numerical magnitude comparison task [$t(72) = 2.04$, $p = 0.046$, $r = 0.23$].² In contrast, RT in the non-symbolic numerical magnitude comparison task did not differ significantly between the two groups [$t(72) = 1.49$, $p = 0.141$, $r = 0.17$]. Moreover, none of the three measures used to evaluate performance in the enumeration of sets in the subitizing range showed significant group differences [mean RT: $t(72) = -1.79$, $p = 0.077$, $r = -0.20$; mean ER: $t(36) = 1.78$, $p = 0.083$, $r = 0.20$, RT slopes: $t(72) = 0.56$, $p = 0.575$, $r = 0.07$]. There was also no significant group difference regarding ER [$t(72) = 0.12$,

TABLE 1 | Comparison of Chinese and German children.

	Chinese children			German children			<i>p</i> (two-sided)
	M	SD	SE	M	SD	SE	
Age	59	7.29	1.20	60	6.99	1.20	$p = 0.90$
Reasoning	19	6.28	1.03	17	4.40	0.72	$p = 0.17$
Processing speed	689	224	37	652	144	24	$p = 0.39$
Counting	27	6.42	1.06	23	6.22	1.02	$p < 0.01$
RT comparison	1,373	424	70	1,513	380	62	$p = 0.14$
ER comparison	32.77	11.46	1.88	37.61	8.84	1.45	$p < 0.05$
RT enumeration 1–3	1,678	430	71	1,522	309	51	$p = 0.08$
ER enumeration 1–3	0	0	0	0.01	0.05	0.01	$p = 0.08$
RT slope 1–3	161	208	34	190	246	40	$p = 0.58$
ER enumeration 4–9	31.98	30.40	5.00	32.88	23.73	3.90	$p = 0.89$
ER slope 4–9	4.94	7.30	1.20	8.26	9.27	1.52	$p = 0.09$

*Descriptive statistics and *p* from two-sample *t*-tests comparing Chinese and German participants with regard to age (in months); reasoning abilities; processing speed (in ms); counting skills; mean RT (in ms) and mean ER (in %) in the non-symbolic numerical magnitude comparison task; mean RT (in ms), mean ER (in %), and RT slopes for the enumeration of sets of dots in the subitizing range (1–3); and mean ER (in %) as well as ER slopes for the enumeration of sets of dots beyond the subitizing range (4–9). *n* = 74 (37 Chinese and 37 German children).*

$p = 0.887$, $r = 0.02$] and ER slopes [$t(68) = 1.71$, $p = 0.091$, $r = 0.19$] for the enumeration of sets of dots beyond the subitizing range. **Table 1** displays an overview of these results. As Levene's test indicated unequal variances for reasoning ($F = 4.82$, $p = 0.031$), mean ER for the enumeration of sets in the subitizing range ($F = 15.28$, $p < 0.001$), and ER slopes for the enumeration

²Performance in trials with ratios close to 1 (i.e., 26/38 vs. 32) was nearly at the chance level of 50% [Chinese children: ER = 45%, German children: ER = 47%]. In order to rule out that the reported group difference is not exclusively due to performance differences in these trials, we compared mean ER only for trials with the other ratios used (i.e., 20/44 vs. 32 and 14/50 vs. 32). In line with the results based on all trials, there was a significant group difference: Chinese children answered more accurately than German children [Chinese children: ER = 27%, German children: ER = 33%, $t(72) = 2.11$, $p = 0.039$, $r = 0.24$].

of sets beyond the subitizing range ($F = 4.09, p = 0.047$), degrees of freedom were adjusted.

In *post hoc* analyses, Pearson correlation coefficients were employed to examine associations between ER in the non-symbolic numerical magnitude comparison task and counting skills in both groups. No significant correlation was found for both Chinese ($r = -0.001, p = 0.994$) and German children ($r = -0.30, p = 0.072$). Using the Fisher r -to- z transformation to compare the correlation coefficients of both groups directly did not reveal a significant difference ($r = -0.001$ vs. $r = -0.30; p = 0.203$).

DISCUSSION

We compared Chinese and German preschool children regarding their performance in a counting task as well as in a non-symbolic numerical magnitude comparison task assessing their ANS and in an enumeration task assessing their OTS. Chinese children showed better performance in the counting task, which is in agreement with previous findings (e.g., Miller et al., 1995). This superior counting performance was accompanied by a better performance in the non-symbolic numerical magnitude comparison task: Chinese children were more accurate in comparing two visually presented dot arrays with respect to their quantity, while showing similarly short response times as German children. Thus, Chinese preschool children were not only able to count higher, but also showed a better performance in a task assessing the ANS. These performance differences cannot be ascribed to differences in general cognitive abilities as Chinese and German children showed similar reasoning abilities and a similar processing speed.

Group differences with regard to the OTS were statistically not significant. Although there was a trend toward fewer errors in Chinese compared to German children during the enumeration of sets of dots in the subitizing range, there was also a trend toward longer reaction times in Chinese children. Similarly, there were no significant group differences with regard to the enumeration of sets of dots beyond the subitizing range (4–9). There was, however, a trend toward a steeper error rate slope in German compared to Chinese children. This might be seen as a further indication of better counting skills of Chinese children. This interpretation must, however, be taken with caution because enumerating sets of dots beyond subitizing range may not only involve counting but also other processes like approximate estimation (Piazza, 2010). Most importantly, the findings of this study reveal that there is no clear indication of advantages for Chinese children in terms of enumerating small sets of items in the subitizing range.

In accordance with previous findings by Rodic et al. (2014), the observed advantage of Chinese children in the non-symbolic numerical magnitude comparison task is statistically significant, but the associated effect size is small ($r = 0.23$). Rodic et al. (2014) assumed a relatively small influence of the ANS on the acquisition of mathematical skills. Meta-analytic findings by Chen and Li (2014) support this view by revealing a small but significant correlation ($r = 0.20$) between the performance in non-symbolic numerical magnitude comparison tasks and mathematical skills. In the present study, no significant correlation between children's

performance in the non-symbolic numerical magnitude comparison task and their counting skills could be observed in both groups. A possible reason for this finding might be that asking children to recite the number word sequence from 1 to 30 is not comprehensive enough to be used as a measure of their early mathematical skills.

Rodic et al. (2014) additionally suggested that mathematical learning affects the ANS. In line with this view, the meta-analysis by Chen and Li (2014) revealed that while non-symbolic numerical magnitude processing skills predict later math performance ($r = 0.24$, based on six longitudinal samples), they can also be predicted by earlier math performance ($r = 0.17$, based on five longitudinal samples). Non-symbolic numerical magnitude processing skills may thus be reciprocally related to mathematical learning. Consequently, the present findings may be explained by two possible underlying mechanisms—on the one hand, more precise ANS representations may enable Chinese children to develop more elaborate counting skills than their German peers. More precise ANS representations of Chinese children might be traced back to more sophisticated visual-spatial skills (Zhou et al., 2015; see also Lonnemann et al., 2019), which have been observed as early as in preschool age and which are assumed to be a consequence of learning to read Chinese characters (see McBride-Chang et al., 2011; McBride and Wang, 2015). In this regard, it has been suggested that performance in visually presented non-symbolic numerical magnitude comparison tasks depends on the ability to integrate different visual cues (Gebuis et al., 2016). On the other hand, Chinese children's more elaborate counting skills may result in more precise ANS representations. This assumption is corroborated by findings showing better counting skills in Chinese children than in Finnish children irrespective of age, but better performance in relational skills in Chinese children only among older children (see Aunio et al., 2006). Aunio et al. (2006) assumed that Chinese children's relative gain in relational skills is a result of the more systematic teaching of counting skills in China. Similarly, more systematic teaching and the associated higher experience and familiarity with counting among Chinese children could have led to better non-symbolic numerical magnitude processing skills compared to German children. Longitudinal studies are needed to further explore this issue. By assessing both the development of non-symbolic numerical magnitude processing skills and the development of counting skills in Chinese and German children over a longer period of time, we would gain a better understanding of the interrelationship between these skills. Moreover, it would be possible to examine whether the direction of influence changes in the course of development and to determine to what extent the developmental trajectories are culture-specific. It can, however, not be ruled out that other factors also play a role. For example, the more regular and transparent Chinese number word system may explain Chinese children's advantage in the counting tasks (see, e.g., Ng and Rao, 2010). If Chinese and German participants attempted to count the dots presented in the non-symbolic numerical magnitude comparison task, differences in the structure of the number naming systems may explain Chinese children's advantage in this task. Indeed, it could be argued that children tried to count the dots in the non-symbolic numerical magnitude comparison task

and that the more accurate performance of Chinese children in this task is merely due to their superior counting skills. In previous studies examining preschool children's ANS, short presentation times were used in the non-symbolic numerical magnitude comparison task in order to prevent children from using counting strategies (see e.g., Libertus et al., 2011, 2013; Mazzocco et al., 2011). In the present study, the sets of different numerical quantities were presented up to a maximum duration of 6,000 ms. Indeed, Libertus et al. (2011, 2013) as well as Mazzocco et al. (2011) used shorter presentation times but they also used smaller set sizes (Libertus et al.: 4–15; Mazzocco et al.: 1–14) which may be more likely to trigger the use of counting strategies. Nevertheless, we cannot rule out that some of our participants attempted to count some of the stimuli. However, the instruction to not to count the dots, the number of dots (14–50), and the restricted response time (6,000 ms) should have prevented this strategy in our study. Moreover, the distribution of reaction times in our study (Chinese children: $M = 1,373$ ms, $SD = 424$; German children: $M = 1,513$ ms, $SD = 380$) indicates that the participants generally identified the side of the larger numerical magnitude without using counting strategies.

With regard to reaction times in the non-symbolic numerical magnitude comparison task, it has to be noted that we unexpectedly observed no significant effect of ratio in Chinese children and a reversed effect of ratio in German children. This indicates that reaction times in non-symbolic numerical magnitude comparison tasks cannot be considered a reliable indicator of the ANS, at least in preschool children. In this regard, it has been demonstrated that accuracy/ER-based measures are more informative about the underlying ANS acuity than RT-based measures (see Dietrich et al., 2016). In addition, recent meta-analyses revealed higher correlations between non-symbolic numerical magnitude processing skills and symbolic math performance for overall accuracy/ER compared to overall RT in a non-symbolic numerical magnitude processing task (Fazio et al., 2014; Schneider et al., 2017).

The superior counting performance of Chinese children was not accompanied by a better performance of Chinese children in

enumerating small sets of items in the subitizing range. This finding does not exclude a contribution of the OTS to the acquisition of counting skills. Indeed, the OTS might be a necessary condition for the acquisition of counting skills, but it does not seem to be related to the observed difference between Chinese and German children's counting skills. Our findings also suggest that the OTS is not affected by the development of counting skills. Due to the small sample size, the results of our study must, however, be viewed with caution. Future studies are thus needed to substantiate our findings and the aforementioned suggestions.

To conclude, results from our study revealed that differences in counting performance between Chinese and German preschool children are accompanied by differences in a non-symbolic numerical magnitude comparison task used to assess the ANS, but not by differences in an enumeration task used to evaluate the OTS. A superior counting performance of Chinese children was thus found to be reflected in the ANS but not in the OTS.

AUTHOR CONTRIBUTIONS

JLo, SLi, JLi, SLin, MH, and SY substantially contributed to the conception and design of the work. JLo and PZ contributed to the acquisition and analysis of data. JLo, SLi, PZ, JLi, SLin, MH, and SY substantially contributed to the interpretation of data for the work and to drafting the work and revising it critically for important intellectual content. JLo, SLi, PZ, JLi, SLin, MH, and SY agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

This research was funded by the Hessian initiative for the development of scientific and economic excellence (LOEWE).

REFERENCES

- Aunio, P., Aubrey, C., Godfrey, R., Pan, Y., and Liu, Y. (2008). Children's early numeracy in England, Finland and People's Republic of China. *Int. J. Early Years Educ.* 16, 203–221. doi: 10.1080/09669760802343881
- Aunio, P., Niemivirta, M., Hautamäki, J., Van Luit, J. E., Shi, J., and Zhang, M. (2006). Young children's number sense in China and Finland. *Scand. J. Educ. Res.* 50, 483–502. doi: 10.1080/00313830600953576
- Bultheller, S., and Häcker, H. O. (2002). *Coloured Progressive Matrices (CPM). Deutsche Bearbeitung und Normierung nach J. C. Raven*. Frankfurt: Pearson Assessment.
- Chen, Q., and Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychol.* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Dietrich, J. F., Huber, S., Klein, E., Willmes, K., Pixner, S., and Moeller, K. (2016). A systematic investigation of accuracy and response time based measures used to index ANS acuity. *PLoS One* 11:e0163076. doi: 10.1371/journal.pone.0163076
- Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53–72. doi: 10.1016/j.jecp.2014.01.013
- Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002
- Gebuis, T., Kadosh, R. C., and Gevers, W. (2016). Sensory-integration system rather than approximate number system underlies numerosity processing: a critical review. *Acta Psychol.* 171, 17–35. doi: 10.1016/j.actpsy.2016.09.003
- Gebuis, T., and Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behav. Res. Methods* 43, 981–986. doi: 10.3758/s13428-011-0097-5
- Gray, S. A., and Reeve, R. A. (2014). Preschoolers' dot enumeration abilities are markers of their arithmetic competence. *PLoS One* 9:e94428. doi: 10.1371/journal.pone.0094428
- Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *Proc. Natl. Acad. Sci. USA* 109, 11116–11120. doi: 10.1073/pnas.1200196109
- Inglis, M., and Gilmore, C. (2014). Indexing the approximate number system. *Acta Psychol.* 145, 147–155. doi: 10.1016/j.actpsy.2013.11.009
- Izard, V., Sann, C., Spelke, E. S., and Streri, A. (2009). Newborn infants perceive abstract numbers. *Proc. Natl. Acad. Sci. USA* 106, 10382–10385. doi: 10.1073/pnas.0812142106
- Libertus, M. E., Feigenson, L., and Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Dev. Sci.* 14, 1292–1300. doi: 10.1111/j.1467-7687.2011.01080.x

- Libertus, M. E., Feigenson, L., and Halberda, J. (2013). Is approximate number precision a stable predictor of math ability?. *Learn. Individ. Differ.* 25, 126–133. doi: 10.1016/j.lindif.2013.02.001
- Lonnemann, J., Müller, C., Büttner, G., and Hasselhorn, M. (2019). The influence of visual-spatial skills on the association between processing of non-symbolic numerical magnitude and number word sequence skills. *J. Exp. Child Psychol.* 178, 184–197. doi: 10.1016/j.jecp.2018.09.018
- Major, C. S., Paul, J. M., and Reeve, R. A. (2017). TEMA and dot enumeration profiles predict mental addition problem solving speed longitudinally. *Front. Psychol.* 8, 2263. doi: 10.3389/fpsyg.2017.02263
- Mazzocco, M. M. M., Feigenson, L., and Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS One* 6:e23749. doi: 10.1371/journal.pone.0023749
- McBride, C., and Wang, Y. (2015). Learning to read Chinese: universal and unique cognitive cores. *Child Dev. Perspect.* 9, 196–200. doi: 10.1111/cdep.12132
- McBride-Chang, C., Zhou, Y., Cho, J.-R., Aram, D., Levin, I., and Tolchinsky, L. (2011). Visual spatial skill: a consequence of learning to read?. *J. Exp. Child Psychol.* 109, 256–262. doi: 10.1016/j.jecp.2010.12.003
- Miller, K. F., Smith, C. M., Zhu, J., and Zhang, H. (1995). Preschool origins of cross-national differences in mathematical competence: the role of number-naming systems. *Psychol. Sci.* 6, 56–60.
- Mullis, I. V. S., Martin, M. O., Foy, P., and Arora, A. (2012). *TIMSS 2011 international results in mathematics*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center. doi: 10.1159/000341768
- Ng, S. S. N., and Rao, N. (2010). Chinese number words, culture, and mathematics learning. *Rev. Educ. Res.* 80, 180–206. doi: 10.3102/0034654310364764
- OECD (2013). *PISA 2012 results: what students know and can do—student performance in mathematics, reading and science*, Vol. I. PISA: OECD Publishing.
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends Cogn. Sci.* 14, 542–551. doi: 10.1016/j.tics.2010.09.008
- Pylyshyn, Z. (2001). Visual indexes, preconceptual objects, and situated vision. *Cognition* 80, 127–158. doi: 10.1016/S0010-0277(00)00156-6
- Reeve, R., Reynolds, F., Humberstone, J., and Butterworth, B. (2012). Stability and change in markers of core numerical competencies. *J. Exp. Psychol. Gen.* 141, 649–666. doi: 10.1037/a0027520
- Revkin, S. K., Piazza, M., Izard, V., Cohen, L., and Dehaene, S. (2008). Does subitizing reflect numerical estimation?. *Psychol. Sci.* 19, 607–614. doi: 10.1111/j.1467-9280.2008.02130.x
- Rodic, M., Zhou, X., Tikhomirova, T., Wei, W., Malykh, S., Ismatulina, V., et al. (2014). Cross-cultural investigation into cognitive underpinnings of individual differences in early arithmetic. *Dev. Sci.* 18, 165–174. doi: 10.1111/desc.12204
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372
- Van Luit, J. E. H., van de Rijt, B. A. M., and Pennings, A. H. (1994). *Utrechtse Getalbegrip Toets [Utrecht test of number sense]*. Doetinchem, The Netherlands: Graviant.
- van Oeffelen, M. P., and Vos, P. G. (1982). A probabilistic model for the discrimination of visual number. *Percept. Psychophys.* 32, 163–170. doi: 10.3758/BF03204275
- Wang, J., and Lin, E. (2009). A meta-analysis of comparative studies on Chinese and US students' mathematics performance: implications for mathematics education reform and research. *Educ. Res. Rev.* 4, 177–195. doi: 10.1016/j.edurev.2009.06.003
- Wang, J., and Lin, E. (2013). Corrigendum to a meta-analysis of comparative studies on Chinese and US student mathematics performance: implications for mathematics education reform and research [Edu. Res. Rev. 4 (3)(2009) 177–195]. *Educ. Res. Rev.* 10, 150–157. doi: 10.1016/j.edurev.2013.10.003
- Zhou, X., Wei, W., Zhang, Y., Cui, J., and Chen, C. (2015). Visual perception can account for the close relation between numerosity processing and computational fluency. *Front. Psychol.* 6:1364. doi: 10.3389/fpsyg.2015.01364

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Lonnemann, Li, Zhao, Linkersdörfer, Lindberg, Hasselhorn and Yan. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Numerical Processing Impairment in 22q11.2 (LCR22-4 to LCR22-5) Microdeletion: A Cognitive-Neuropsychological Case Study

Lívia de Fátima Silva Oliveira^{1,2}, Annelise Júlio-Costa¹, Fernanda Caroline dos Santos³, Maria Raquel Santos Carvalho^{4*} and Vitor Geraldí Haase^{1,2,5,6,7}

¹ Laboratório de Neuropsicologia do Desenvolvimento, Departamento de Psicologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ² Programa de Pós-Graduação em Neurociências, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ³ Programa de Pós-Graduação em Genética, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁴ Departamento de Biologia Geral, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁵ Programa de Pós-graduação em Psicologia, Cognição e Comportamento, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁶ Programa de Pós-graduação em Saúde da Criança e do Adolescente, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ⁷ Instituto Nacional de Ciência e Tecnologia sobre Comportamento, Cognição e Ensino, São Carlos, Brazil

OPEN ACCESS

Edited by:

Marcus Lindskog,
Uppsala University, Sweden

Reviewed by:

Kevin Antshel,
Syracuse University, United States
Ann Dowker,
University of Oxford, United Kingdom

*Correspondence:

Maria Raquel Santos Carvalho
mraquel@icb.ufmg.br

Specialty section:

This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

Received: 01 May 2018

Accepted: 23 October 2018

Published: 21 November 2018

Citation:

Oliveira LFS, Júlio-Costa A, Santos FC, Carvalho MRS and Haase VG (2018) Numerical Processing Impairment in 22q11.2 (LCR22-4 to LCR22-5) Microdeletion: A Cognitive-Neuropsychological Case Study. *Front. Psychol.* 9:2193. doi: 10.3389/fpsyg.2018.02193

Although progress has been made, the cognitive, biological and, particularly, the genetic underpinnings of math learning difficulties (MD) remain largely unknown. This difficulty stems from the heterogeneity of MD and from the large contribution of environmental factors to its etiology. Understanding endophenotypes, e.g., the role of the Approximate Number System (ANS), may help understanding the nature of MD. MD associated with ANS impairments has been described in some genetic conditions, e.g., 22q11.2 deletion syndrome (22q11.2DS or Velocardiofacial syndrome, VCFS). Recently, a girl with MD was identified in a school population screening. She has a new syndrome resulting from a microdeletion in 22q11.2 (LCR22-4 to LCR22-5), a region adjacent to but not overlapping with region 22q11.2 (LCR22-2 to LCR22-4), typically deleted in VCFS. Here, we describe her cognitive-neuropsychological and numerical-cognitive profiles. The girl was assessed twice, at 8 and 11 years. Her numerical-cognitive performance at both times was compared to demographically similar girls with normal intelligence in a single-case, quasi-experimental study. Neuropsychological assessment was normal, except for relatively minor impairments in executive functions. She presented severe and persistent difficulties in the simplest single-digit calculations. Difficulties in commutative operations improved from the first to the second assessment. Difficulties in subtraction persisted and were severe. No difficulties were observed in Arabic number writing. Difficulties in single-digit calculation co-occurred with basic numerical processing impairments in symbolic and non-symbolic (single-digit comparison, dot sets size comparison and estimation) tasks. Her difficulties suggest ANS impairment. No difficulties were detected in visuospatial/visuoconstructional and in phonological processing tasks. The main contributions of the present study are: (a) this is the first characterization of the neuropsychological phenotype in 22q11.2DS (LCR22-4 to LCR22.5) with normal intelligence; (b) mild forms of specific genetic conditions contribute to persistent MD

in otherwise typical persons; (c) heterogeneity of neurogenetic underpinnings of MD is suggested by poor performance in non-symbolic numerical processing, dissociated from visuospatial/visuoconstructional and phonological impairments; (d) similar to what happens in 22q11.2DS (LCR22-2 to LCR22-4), ANS impairments may also characterize 22q11.2DS (LCR22-4 to LCR22-5).

Keywords: math learning difficulties, developmental dyscalculia, 22q11.2DS (LCR22-4 to LCR22-5), cognitive phenotype, Weber fraction, approximate number system

INTRODUCTION

Number processing abilities, such as magnitude comparison and estimation, and knowledge about the numerals, have been implicated in both typical and atypical math learning (Siegler and Braithwaite, 2017). Current discussions focus on the role of accuracy of numerical magnitude representations in the non-symbolic Approximate Number System (ANS) vs. access to these non-symbolic representations through symbolic numbers (Leibovich et al., 2017).

Accuracy in non-symbolic numerical representations has been linked to both typical (Halberda et al., 2008) and atypical (Landerl et al., 2004; Piazza et al., 2010; Mazzocco et al., 2011; Pinheiro-Chagas et al., 2014) math learning. Other studies suggest that symbolic numerical representations play a more important role (Rousselle and Noël, 2007; De Smedt and Gilmore, 2011; Szucs et al., 2013, see review in De Smedt et al., 2013). Meta-analyses indicate that correlations between number processing and math achievement are weak (r 's between 0.2 and 0.3) and are slightly larger for symbolic over non-symbolic numerical processing (Chen and Li, 2014; Fazio et al., 2014; Schneider et al., 2017). It is also unknown how and when non-symbolic and symbolic processing influence math learning (Leibovich et al., 2017).

Developmental dyscalculia, math learning difficulties (MD) and number processing impairments have been described for some syndromes of environmental or genetic etiology such as fetal alcohol syndrome (Jacobson et al., 2011), fragile X syndrome in females (Mazzocco, 2001; Villalon-Reina et al., 2013), Turner's syndrome (Bruandet et al., 2004; Zougkou and Temple, 2016), Williams-Beuren syndrome (Krajcsi et al., 2009; Libertus et al., 2014) and velocardiofacial syndrome (VCFS, 22q11.2 deletion syndrome, 22q11.2DS) (Barnea-Goraly et al., 2005; De Smedt et al., 2009; Attout et al., 2017).

The presence of developmental dyscalculia among the phenotypes of many different genetic syndromes suggests that multiple specific genetic factors contribute to the emergence of dyscalculia. As the genotype-phenotype variability of genetic syndromes is large, milder forms of a given syndrome may eventually contribute to MD, particularly in individuals with normal intelligence.

One of the most investigated syndromes associated with developmental dyscalculia is 22q11.2DS, resulting from microdeletions of a specific region on chromosome 22. Chromosome 22 is the second smallest human chromosome and corresponds to approximately 1.6% of human genomic DNA

(Genome Reference Consortium, 2018). Genetic alterations on chromosome 22q11.2 have been associated with numerous health conditions, including intellectual disability and schizophrenia. At least 48 genes have been identified in the region associated with 22q11.2DS, including *PRODH* and *COMT*, implicated in cognitive functions through regulation of dopamine metabolism (Karayiorgou et al., 2010; Espe, 2018).

The long arm of chromosome 22 contains interspaced low copy-repeated (LCR) sequences, which make this region susceptible to non-homologous recombination events leading to microdeletions or microduplications. Persons having typical 22q11.2DS present the microdeletion of the 22q11.2 (LCR22-2 to LCR22-4) interval.

To elucidate the genomic variations contributing to math learning difficulties, in a previous population study ($n = 1,520$ children), we investigated some genotypic and phenotypic characteristics of MD children, defined as standardized math achievement below the PR 25 (Ferreira et al., 2012). Among 82 MD children, we identified a 8-year-old girl presenting a microdeletion on chromosome 22q11.2 in the LCR22-4 to LCR22-5 interval (Carvalho et al., 2014).

Reviewing the literature, Carvalho and coworkers characterized a new syndrome, 22q11.2DS (LCR22-4 to LCR22-5), associated with microdeletions spanning only this interval and not extending proximally into the 22q11.2 (LCR22-2 to LCR22-4) interval (typically deleted in 22q11.2DS) or distally, into the 22q11.2 (LCR22-5 to LCR22-6) interval. Further, the authors proposed 22q11.2DS (LCR22-4 to LCR22-5) as an additional cause of dyscalculia in 22q11.2.

22q11.2DS (LCR22-4 to LCR22-5) is characterized by intellectual disability in most cases, and psychiatric symptoms and MD suggesting a heterogeneous condition (Table 1). To date, neither the neuropsychological phenotype nor the impairments in number processing have been detailed. Here, we describe a single-case, quasi-experimental study developed to characterize the cognitive-neuropsychological and numerical-cognitive endophenotypes underlying math learning difficulties in the child having the 22q11.2DS (LCR22-4 to LCR22-5) described by Carvalho et al. (2014).

22q11.2DS (LCR22-4 to LCR22-5) has already been reported in 33 persons (Saitta et al., 1999; Mikhail et al., 2007, 2014; Ben-Shachar et al., 2008; Newbern et al., 2008; Rodningen et al., 2008; Xu et al., 2008; Beaujard et al., 2009; Bruce et al., 2010; Tan et al., 2011; Verhoeven et al., 2011; Fagerberg et al., 2012; Molck et al., 2013; Carvalho et al., 2014; Lindgren et al., 2015; Spineli-Silva et al., 2017).

TABLE 1 | Findings in patients with 22q11.2DS spanning exclusively the interval LCR22-4 to LCR22-5.

Studies	Sex	Age (year)	Gestational alterations	Postnatal alterations	Physical malformations	Cognitive phenotype	Behavior problems	Specific learning disability
Saitta et al., 1999	M	2	Prematurity	Normal motor development; speech delay; short stature	Cardiac, velopalatine, bone, facial asymmetry	–	–	–
Mikhail et al., 2007	M	15	Prematurity	No development delay	Bone, facial asymmetry	Inferior visual-motor integration (8.4 years). Intellectual disability	Attention deficit hyperactivity disorder (ADHD)	–
Ben-Shachar et al., 2008	M	6	Prematurity	Yes	Cardiac, facial asymmetry, celiac disease	–	No	–
	M	5	Prematurity	No	Facial asymmetry	–	Uncontrolled aggression	–
	M	11	Prematurity	Yes	Velopalatine, bone, facial asymmetry, obesity, karyotype 47,XXX	–	Yes	–
	M	3	Prematurity	Yes	Cardiac, velopalatine, facial asymmetry	–	No	–
	F	3	Prematurity	Yes	Facial asymmetry	–	–	–
	M	4	Prematurity	Yes	Velopalatine, bone, facial asymmetry	–	–	–
Newbern et al., 2008	F	–	–	Restricted postnatal growth	Cardiac, facial asymmetry	Intellectual disability	–	–
	M	–	–	Restricted postnatal growth	Cardiac, velopalatine, bone, facial asymmetry	Intellectual disability	–	–
Rodningen et al., 2008	F	7	Prematurity	Mild psychomotor delay; low muscle tone	Cardiac, bone, facial asymmetry	–	–	–
	M	7	Prematurity	Speech delay	Velopalatine, bone, facial asymmetry	Difficulties in: language comprehension, articulate some sounds, motor tasks	Cooperative person, but he challenges limits set by his parents; good in keep the routines.	–
Xu et al., 2008	M	11 months	Prematurity	–	Cardiac, velopalatine, facial asymmetry	Functioning at a 6–7 months level	–	–
Beaujard et al., 2009	F	35	–	–	Cardiac, velopalatine, facial asymmetry	Intellectual disability	–	–
	M	2 months	Prematurity	–	Cardiac, facial asymmetry	–	–	–
Bruce et al., 2010	F	12	Prematurity	Postnatal growth, motor delay	Cardiac, velopalatine, bone, facial asymmetry	–	–	–
Tan et al., 2011	F	–	Prematurity	Hypotonia	Cardiac, bone, facial asymmetry	–	No	–
Verhoeven et al., 2011	F	18	Prematurity	Psychomotor delay; eating problems	Cardiac, velopalatine, bone, facial asymmetry	Difficulties: planning; concentration; visuospatial perception	Impulsivity mood instability, anxiety; paranoid ideation	Yes. In calculation.

(Continued)

TABLE 1 | Continued

Studies	Sex	Age (year)	Gestational alterations	Postnatal alterations	Physical malformations	Cognitive phenotype	Behavior problems	Specific learning disability
Fagerberg et al., 2012	F	14	Prematurity	–	Cardiac, facial asymmetry	–	Attention deficit	–
	M	13	Prematurity	Global developmental delay	Velopalatine, bone, facial asymmetry	–	–	–
Molck et al., 2013	F	4	Prematurity	Developmental delay: motor delay	Cardiac, bone, facial asymmetry	–	Agitation and attention deficit	–
	F	4	–	Developmental delay	Velopalatine, bone, facial asymmetry	–	–	–
Carvalho et al., 2014 (this study)	F	11	–	–	Velopalatine, bone, facial asymmetry	Normal intelligence; mild executive function deficits	Difficulties in social interaction (social phobia). ADHD	Yes, especially in math and language comprehension
Mikhail et al., 2014	F	7,8	Prematurity	Global developmental delay	Bone, facial asymmetry	Intellectual disability	Talks to oneself and to imaginary friends; social, immaturity	–
	M	9	Prematurity	Global developmental delay	Cardiac, velopalatine, bone, facial asymmetry, seizures	Intellectual disability	Poor impulse control and anger issues	–
	M	20	–	Global developmental delay	Bone, facial asymmetry, seizures, pituitary tumor	Intellectual disability	–	–
	F	10	Prematurity	Global Developmental delay	Cardiac, facial asymmetry	Intellectual disability	ADHD; social immaturity, anxiety; impulsivity	–
	F	20	Prematurity	Global Developmental delay	Seizures	Intellectual disability	ADHD; Asperger Disorder	–
	M	22	Prematurity	Global Developmental delay	Cardiac, facial asymmetry	Intellectual disability	–	–
Lindgren et al., 2015	F	21 months	Prematurity	Growth restriction, motor delay	Cardiac, bone, facial asymmetry	–	–	–
	F	5	Prematurity	Speech and language, speech often unintelligible	Bone, facial asymmetry	Learning disorder, individualized educational program	ADHD, pediatric bipolar disorder, aggressive behavior	–
	F (sister)	4	Prematurity	Speech and language, unintelligible speech,	Velopalatine, facial asymmetry	Deficit in visual perception and motor integration, mildly delayed gross motor milestones. Individualized educational program in place	ADHD by DSM-V Oppositional defiant disorder/conduct disorder, aggressive behavior	–
Spineli-Silva et al., 2017	F	11	Prematurity	Speech and developmental delay	Cardiac, velopalatine, bone, facial asymmetry	Intellectual disability	–	–

In general, the published studies describe in broad strokes the phenotypic and genotypic characteristics related to 22q11.2DS (LCR22-4 to LCR22-5) (Table 1), which can be summarized in five topics:

- 1) Distal microdeletions are a health condition independent of 22q11.2DS. Deletions and duplications in the 22q11.2 region are classified as proximal, central and distal (types I, II and III) (Burnside, 2015). 22q11.2DS (LCR22-4 to LCR22-5) is considered distal type I;
- 2) Although there is no consensus whether distal microdeletions cause a more subtle (Saitta et al., 1999; Mikhail et al., 2007; Carvalho et al., 2014) or more severe phenotype (Ben-Shachar et al., 2008; Xu et al., 2008; Tan et al., 2011; Lindgren et al., 2015), some characteristics have been described more frequently: (a) congenital heart diseases (most frequently of the Truncus Arteriosus type) observed in 16 of the 32 persons reported in the literature (Saitta et al., 1999; Mikhail et al., 2007, 2014; Ben-Shachar et al., 2008; Newbern et al., 2008; Xu et al., 2008; Beaujard et al., 2009; Bruce et al., 2010; Tan et al., 2011; Verhoeven et al., 2011; Fagerberg et al., 2012; Molck et al., 2013; Spineli-Silva et al., 2017); (b) prematurity and low birth weight reported in almost all patients (Saitta et al., 1999; Mikhail et al., 2007, 2014; Ben-Shachar et al., 2008; Newbern et al., 2008; Rodningen et al., 2008; Xu et al., 2008; Beaujard et al., 2009; Bruce et al., 2010; Tan et al., 2011; Verhoeven et al., 2011; Fagerberg et al., 2012; Molck et al., 2013; Lindgren et al., 2015; Spineli-Silva et al., 2017); (c) language development delay observed in six patients (Saitta et al., 1999; Ben-Shachar et al., 2008; Rodningen et al., 2008; Fagerberg et al., 2012; Lindgren et al., 2015; Spineli-Silva et al., 2017); d) bone malformations reported in 15 patients (Saitta et al., 1999; Mikhail et al., 2007, 2014; Ben-Shachar et al., 2008; Rodningen et al., 2008; Bruce et al., 2010; Tan et al., 2011; Fagerberg et al., 2012; Molck et al., 2013; Spineli-Silva et al., 2017); and, e) facial dysmorphisms marked by micrognathia, microcephaly, narrow palpebral fissures, arched eyebrows, featureless filter, hypertelorism, prominent nose, pointed chin, thin lips, etc. One or more of these phenotypes are reported in at least one patient (Saitta et al., 1999; Mikhail et al., 2007, 2014; Ben-Shachar et al., 2008; Newbern et al., 2008; Rodningen et al., 2008; Xu et al., 2008; Beaujard et al., 2009; Bruce et al., 2010; Tan et al., 2011; Verhoeven et al., 2011; Fagerberg et al., 2012; Molck et al., 2013; Lindgren et al., 2015; Spineli-Silva et al., 2017);
- 3) There is heterogeneity in intelligence. Most studies have qualitatively characterized intellectual disability. Normal or borderline intelligence is described for some patients (Ben-Shachar et al., 2008; Verhoeven et al., 2011; Carvalho et al., 2014);
- 4) Behavioral symptoms are briefly cited in most studies: (a) anxiety, social immaturity and social phobia; (b) Attention Deficit Hyperactivity Disorder (ADHD), poor impulse control, anger issues and aggressive behaviors; and,

(c) Asperger disorder (Mikhail et al., 2007, 2014; Ben-Shachar et al., 2008; Carvalho et al., 2014; Lindgren et al., 2015);

- 5) Learning difficulties in mathematics have been reported in two cases with normal or borderline intelligence (Verhoeven et al., 2011; Carvalho et al., 2014). Additionally, Beaujard et al. (2009) described a case with family recurrence in which the mother had a history of learning difficulties.

As mentioned above, developmental dyscalculia is a heterogeneous condition, probably characterized by different subtypes and underlying cognitive mechanisms (Wilson and Dehaene, 2007; Rubinsten and Henik, 2009; Karagiannakis et al., 2014). At least five cognitive mechanisms have been implicated in typical and atypical math achievement: (a) working memory and executive processing, probably associated with ADHD; (b) phonological processing, probably associated with developmental dyslexia; (c) visuospatial and visuoconstructional processing, probably associated with nonverbal learning disability; (d) accuracy of number representations, probably underlying pure cases of developmental dyscalculia; and, eventually, (e) math anxiety, as a compound, aggravating factor.

Number processing deficits in 22q11.2DS (LCR22-4 to LCR22-5) must be contrasted to those observed in typical 22q11.2DS. In the typical 22q11.2DS, two of the most salient cognitive traits associated with developmental dyscalculia are impairments in visuospatial and visuoconstructional processing (Simon et al., 2005a,b; Antshel et al., 2008; Schoch et al., 2014; Wong et al., 2014; Attout et al., 2017), and in the accuracy of non-symbolic and symbolic numerical representations (Simon et al., 2005a,b; De Smedt et al., 2009; Oliveira et al., 2014; Attout et al., 2017; Brankaer et al., 2017). It is not known, for example, whether the numerical and visuospatial processing deficits observed in the typical 22q11.2DS reflect a common underlying impairment or may, eventually, be dissociated. Dissociation between visuospatial and numerical impairments in a case of developmental dyscalculia of genetic origin would be of theoretical interest, and would also hint at the neurobiological systems involved.

So far, no studies have specifically investigated the behavioral and cognitive phenotypes of distal microdeletions in 22q11.2, particularly 22q11.2 (LCR22-4 to LCR22-5). Therefore, the aim of the present study is to investigate and describe in detail the cognitive-neuropsychological characteristics of a girl presenting MD and 22q11.2DS (LCR22-4 to LCR22-5), who was assessed at ages 8 and 11. The underlying assumption is that, although this distal microdeletion is classified as a distinct syndrome, the pattern of general neuropsychological and numerical processing deficits presented by affected persons may resemble that presented by individuals with typical 22q11.2DS. This is based on the observation that some symptoms described for patients having 22q11.2DS (LCR22-4 to LCR22-5) have also been frequently described for patients having typical 22q11.2DS, e.g., conotruncal congenital heart malformations or submucous cleft palate. Therefore, there may be long range effects (Zeititz et al., 2013).

More specifically, we were interested in investigating whether the girl presents impairments in visuospatial and visuoconstructional processing and in the accuracy of numerical representations, two of the most salient phenotypic traits in the typical 22q11.2DS. We were also interested in investigating whether these two forms of impairment are dissociable. To test these hypotheses, we compared her general neuropsychological and numerical-cognitive performance at ages 8 and 11 using a single-case, quasi-experimental design (Crawford et al., 2010).

CLINICAL REPORT

A girl with 22q11.2DS (LCR22-4 to LCR22-5), was identified among children in a population screening for math learning difficulties in Belo Horizonte, Brazil (Ferreira et al., 2012; Oliveira-Ferreira et al., 2012; Carvalho et al., 2014). At the time of the screening, she was 8 years old and attending the 3rd grade of elementary school. Her intelligence was normal and her performance on a standardized arithmetic achievement test was below the PR 25. She was then referred for a comprehensive neuropsychological investigation and genotyping. Results of Multiplex Ligation-dependent Probe Amplification (MLPA) indicated the presence of an atypical distal microdeletion on chromosome 22q11.2. This microdeletion was confirmed, and its size was determined through an array CGH (947,631 bp) (Carvalho et al., 2014).

The girl underwent neuropsychological assessment twice. She was initially assessed at 8 years, by occasion of the population screening, and later at 11 years, when attending the 6th grade. She had shown learning difficulties since the beginning of elementary school. According to her mother, the difficulties had always been more severe in mathematics and in the interpretation of texts. She was retained in the 6th grade because of her math difficulties. This happened at the end of the school year, well after the second neuropsychological assessment. There was no history of difficulties in word reading and spelling or initial literacy acquisition. Her favorite subject at school was English and the girl was able to easily learn song lyrics in English.

The parents described her as a shy girl with a tendency to isolate. Additionally, according to them, the girl used to have problems expressing her needs and exposing her difficulties, especially at school. Her only friend was an 18-year-old cousin. She had difficulties initiating social interactions, especially with peers. Eventually, after becoming acquainted, she would interact normally.

At home, the girl was independent and helped with household chores, but performed at a slow pace and had difficulties concentrating in and finishing chores and homework. She was described as hyperactive, inattentive and anxious. The symptoms of hyperactivity were treated with methylphenidate for 2 months. Treatment was discontinued as the symptoms of inattention remained and anxiety symptoms were exacerbated. She had the habit of nail-biting. Parents reported some minor problems related to aggressive behavior. According to them, the girl would occasionally get into fights with her 6-year-old sister.

No information on pregnancy, delivery or initial development was available, as she was adopted at age 1 year. At that time, she was unable to sit or crawl. After 3 months with the adopted family, she began to walk and to utter her first words. Respiratory problems were constant in the first years of life. The parents also reported that occasionally the girl had nocturnal enuresis up to 7 years and a tendency to withhold urine when playing.

She lived with her adoptive parents and a younger sister, enjoying a stable home environment. The parents married 16 years ago. Both parents completed high school and had no history of learning difficulties. The adoptive father had been employed in the same company for more than 25 years. The adoptive mother was a housewife, who had serious health problems related to systemic lupus erythematosus, requiring constant treatment with corticosteroids. Her younger sister was the biological daughter of the couple. Follow-up disclosed that the biological daughter of the couple presented typical school achievement.

On clinical examination, the girl had short stature, normal weight and head circumference, narrow palpebral fissures, long nose, submucosal cleft of the palate, bifid uvula, pointed chin, long and thin fingers, short and broad nails (Carvalho et al., 2014). Her phenotypic characteristics are organized and compared to other published cases in **Table 1**.

METHODS

The girl participated in a quasi-experimental case study. Her general neuropsychological performance was compared to that of available published Brazilian standards. Numerical-cognitive performance was compared to that of two different but demographically similar groups of typically developing children (Controls) at 8 and 11 years. Typically achieving children participating in the Control group were recruited from public schools and were assessed in the context of the same study in which she was identified. All Controls originated from the same socio-economic background as the girl. Specific statistical procedures were used to compare her performance to that of the Controls (Crawford et al., 2010). At 11 years, she also underwent a psychiatric assessment.

Participants and Procedures

All research procedures complied with the Helsinki principles and were previously approved by the local ethics in research board (Research Ethics Committee of the Federal University of Minas Gerais: CAAE: 0091.0.203.000-10). Informed parental consent was obtained for the purposes of research participation. Informed consent to participate in the study was obtained from the parents in written form and orally from the girl. A specific written consent for publication was also obtained in written form, signed by both the girl and her mother. This informed consent includes their agreement with the publication of the indirectly identifiable information such as gender and age and agreement with the publication of the case report.

All general neuropsychological tests used in the first assessment were reapplied and some tasks were added in the reassessment (**Table 2**). The same battery of numerical-cognitive tasks was used in the two assessments. At 8 years, the girl's

performance in the numerical-cognitive evaluation battery was compared to that of a group of 35 girls (mean age = 8.32 years; $SD = 0.47$ years) attending the 3rd grade of public elementary schools. At 11 years, her performance in the numerical-cognitive evaluation battery was compared to the performance of a group of 24 girls (mean age = 11.38 years; $SD = 0.49$ years) attending the 6th grade of elementary public schools. All the individuals of both Control groups had average intelligence (PRs 50 to 75 on the Raven's Colored Progressive Matrices, Angelini et al., 1999) and did not present learning difficulties as assessed by the TDE Arithmetic and TDE Spelling (Stein, 1994; Oliveira-Ferreira et al., 2012).

Instruments

Behavioral Assessment

At 11 years, her adoptive parents responded the Child Behavior Checklist (CBCL, Rocha et al., 2012), a questionnaire that evaluates behavioral symptoms and psychosocial functioning of individuals aged 6 to 11 years. Her results in the CBCL were compared with the norms for girls of the same age group.

General Neuropsychological Assessment

In **Table 2**, the general neuropsychological domains evaluated when she was 8 and 11 years old, and their respective tasks and normative references, are summarized.

The Brazilian School Achievement Test (TDE), which was used as a criterion of typicality in school achievement, will be discussed in more detail. The TDE is a standardized test of school performance in children from the 2nd to 7th grades. It comprises three subtests, respectively, of Arithmetic, Reading and Spelling (Stein, 1994; Ferreira et al., 2012). The Arithmetic subtest is composed of three simple verbally presented word problems (i.e., Which is the largest, 28 or 42?) and 35 written arithmetic calculations of increasing complexity (i.e., very easy: 4-1; easy: 1230+150+1620; intermediate: 823*96; hard: 3/4+2/8). The single-word Reading subtest of the TDE consists of 70 single-word stimuli, which must be read aloud by the proband. The single-word Spelling subtest consists of dictation of 34 words of increasing syllabic complexity (i.e., *toca*; *balanço*; *cristalização*). The reliability coefficients (Cronbach's α) of the TDE subtests are 0.87 or higher. The TDE has been used in other studies, displaying both reliability and validity in assessing learning difficulties and their cognitive correlates (Moura et al., 2013, 2015; Haase et al., 2014; Lopes-Silva et al., 2014, 2016; Pinheiro-Chagas et al., 2014).

Numerical-Cognitive Assessment

An experimental battery for numerical-cognitive assessment in children and adolescents was used in the present, as well as in previous studies (Costa et al., 2011; Pinheiro-Chagas et al., 2014). The numerical-cognitive battery comprises tasks of number processing and single-digit calculation. The following tasks were used:

Simple Reaction Time (SRT): The computerized RT task is a visual detection task used to control for possible differences in basic processing speed, not related to numerical tasks. In this task the picture of a wolf (height 9.31 cm; length = 11.59 cm) was

displayed in the center of a black screen for a maximum time of 3,000 ms. Participants were instructed to press the spacebar on the keyboard as fast as possible whenever the wolf appeared. Each trial was terminated with the first key press. The task had 30 experimental trials, with an inter-trial interval varying between 2,000 and 8,000 ms. SRT was used to control for eventual effects of general processing speed on the numerical tasks.

Non-symbolic Magnitude Comparison Task: Participants were instructed to compare two sets of black dots, simultaneously presented in two white circles on the left and right sides of the screen. They were required to choose the larger numerosity by pressing a side-congruent key (Pinheiro-Chagas et al., 2014). On each trial, one of the two white circles contained 32 dots (reference numerosity), and the other contained 20, 23, 26, 29, 35, 38, 41, or 44 dots. Each numerosity was presented eight times, and every presentation was arranged in a different spatially pseudo-random configuration. The task comprised 64 testing trials. The maximum stimulus presentation time was 4,000 ms, and the intertrial interval was 700 ms. Between trials, a 3 cm fixation cross appeared on the screen for 500 ms. Non-numerical cues were prevented by using a MATLAB script to design and generate the sets of dots to represent the non-symbolic numerosities (Dehaene et al., 2005). This script was programmed so that, in half of the trials, dot size remained constant and total dot area covaried positively with the numerosity. In the other half of the trials, total dot area remained constant and dot size covaried negatively with numerosity. Each child's data were trimmed to exclude responses ± 3 SD away from the individual mean RT. The internal Weber fraction (w) was calculated for each child as an indicator of approximate number system (ANS) or number sense acuity, based on the Log-Gaussian model of number representation (Dehaene, 2007), using the methods described by Piazza et al. (2004).

Single-digit Magnitude Comparison Task: In another task, developed by Pinheiro-Chagas et al. (2014), Arabic digits from 1 to 9 were presented on the computer screen (2.12 cm height, 2.12 cm length). The visual angle of the stimuli vertically and horizontally comprised 2.43°. The children were instructed to compare the stimuli with the reference number 5. The digits were presented in white on a black background. A predefined key on the left side of the keyboard should be pressed with the left hand, if the presented digit was less than 5. If the digit was greater than 5, a right key should be pressed with the right hand. The digit 5 was never presented on the computer screen (internal reference). Numerical distances between the stimuli and the reference digit (5) varied from 1 to 4. Each numerical distance was presented the same number of times. Between trials, a fixation point of the same size and color as the stimuli was presented on the screen. The task comprised 80 experimental trials. The maximum stimulus presentation time was 4,000 ms, and the intertrial interval was 700 ms. Dependent measures were mean accuracy and reaction times. A efficiency score P can also be used as a measure of symbolic magnitude processing efficiency, penalizing RT for inaccuracy: $P = RT / (1 + 2ER)$ according to Lyons et al. (2014). In the formula, RT means reaction time and ER stands for error rates, considering reaction time (RT) and errors rates (ER) as measures of performance for each child. ERs

TABLE 2 | Neuropsychological assessment battery.

Domain assessed	Test	References
Intelligence	Raven's Colored Progressive Matrices	Angelini et al., 1999
	Wechsler Intelligence Scale for Children (WISC-III)	Figueiredo, 2002
School achievement	Brazilian School Achievement Test (TDE)	Stein, 1994; Oliveira-Ferreira et al., 2012
Reading-related abilities	Nonword repetition	Santos and Bueno, 2003
	Nonword Reading	Lopes-Silva et al., 2014
	Phoneme elision	Lopes-Silva et al., 2014
Behavior and psychosocial functioning	Child Behavior Checklist (CBCL)	Rocha et al., 2012
Motor dexterity	9-Hole Peg Test (9-HPT)	Poole et al., 2005
Body representation	Finger localization task	Costa et al., 2011
	Right-left orientation test	Costa et al., 2011
Visuospatial/Visuoconstructional abilities	Rey-Osterrieth Complex Figure copy	Oliveira et al., 2004
Episodic memory	Rey Auditory-verbal Learning Test (RAVLT)	Lacerda, 2012
Short-term and working memory	Digit span	Figueiredo and Nascimento, 2007
	Corsi blocks	Santos et al., 2005
	Consonantal trigrams	Vaz et al., 2010
Executive functions	Semantic word fluency	In house
	5-point design fluency test	In house
	Trail Making Test (TMT) A and B	In house
	Victoria Stroop color-word interference test	Charchat-Fichman and Oliveira, 2009
Numerical-cognitive abilities	Simple reaction time	Costa et al., 2011; Ferreira et al., 2012; Pinheiro-Chagas et al., 2014
	Non-symbolic magnitude comparison task	
	Single-digit Magnitude Comparison task	
	Set-size Magnitude Estimation	
	Arabic number reading task	
	Arabic number writing task	
	Single-digit operations	
	Simple Word Problems	

were multiplied by 2 because the task was a binary forced choice ($ER = 0.5$ indicates chance level). Higher scores indicate worse performance. If the performance were perfectly accurate, P would correspond to the individual's average RT ($P = RT$).

Set-size Magnitude Estimation: In the non-symbolic magnitude estimation task, participants were asked to verbally estimate the quantity of dots shown on the computer screen (Pinheiro-Chagas et al., 2014). The stimuli were black dots presented in a white circle over a black background. The numerosities were 10, 16, 24, 32, 48, 56 or 64 dots. Each numerosity was presented 5 times, each time in a different configuration. The same numerosity never appeared in consecutive trials. The task comprised 35 trials. Counting was avoided by setting the maximum stimulus presentation time to 1000 ms. The examiner, who was seated next to the child, pressed the spacebar and entered the child's response as soon as the child responded. A 3-cm wide/long fixation cross appeared on the screen between individual trials. Use of non-numerical cues was prevented by programming the stimuli in the same

manner as those of the non-symbolic number comparison task, described above. Memorization effects due to the repetition of a specific stimulus were avoided in that, in each trial, the stimuli were randomly chosen from a set of 10 precomputed images with the given numerosity. For each subject, data were trimmed to exclude the responses ± 3 SD from the mean chosen value across all of the trials. The mean coefficient of variation (cv) was selected as the dependent measure of ANS-accuracy.

Arabic Number Reading Task: Twenty-eight Arabic numbers printed in a booklet were presented, one at a time, and the child had to read the numbers aloud (Moura et al., 2015). The set of items consisted of numbers with up to 4 digits (3 numbers with one digit, 9 numbers with two digits, 8 numbers with three digits and 8 numbers with four digits). The internal consistency of the task is $KR-20 = 0.90$.

Arabic Number Writing Task: The participant was instructed to write dictated numbers using Arabic numerals (Moura et al., 2015). This task was composed of 40 items, and the numbers contained up to 4 digits (3 numbers with one digit, 9 numbers

with two digits, 10 numbers with three digits and 18 numbers with four digits). The internal consistency of the task is 0.96 with the KR-20 formula.

Single-digit operations: This task consisted of single-digit addition (27 items), subtraction (27 items), and multiplication (28 items) operations for individual application, which were printed on separate sheets of paper (Costa et al., 2011). Children were instructed to answer as fast and as accurately as they could; time limit per block was 1 min. Arithmetic operations were organized in two levels of complexity and were presented to the children in separate blocks: one block consisted of simple arithmetic table facts and the other block of more complex problems. Simple additions were defined as those operations having results below 10 (e.g., $3 + 5$), while complex additions were those having results between 11 and 17 (e.g., $9 + 5$). Tie problems (e.g., $4 + 4$) were not used for addition. Simple subtractions were defined as those operations having operands less than 10 (e.g., $9 - 6$), while complex subtractions were defined as those having operands ranging from 11 to 17 (e.g., $16 - 9$). No negative results were included in the subtraction problems. Simple multiplications were defined as those operations having results less than 25 and/or with the digit 5 as one of the operands (e.g., 2×7 , 5×6), while complex multiplications were defined as those having products ranging from 24 to 72 (6×8). Tie problems were not used for multiplication. Reliability coefficients were high (Cronbach's $\alpha > 0.90$).

Simple Word Problems: Twelve simple arithmetic problems (e.g., "Gabi has 3 reais. Debora has 6 reais. How much do they have together?") were read aloud by the examiner and simultaneously presented in written form. The child had to solve the problem mentally and write the answer on the paper, with a time limit of 1 min per problem. The dependent variable was the number of correct responses (for more details, see Costa et al., 2011).

Statistical Analysis

All scores were z-standardized for age to facilitate comparisons. In the comparison with the published norms, a deviation of 1.5 SD from the mean was used as the cut-off score to determine whether the domain was impaired or preserved. A cut-off score of test performance was employed because diagnosis implies categorization: either the person presents or does not present some health condition. The cut-off score chosen is not overly restrictive or excessively compliant. Larger time executions in the 9-Hole Peg Test (Poole et al., 2005), Trail Making Test, and Victoria Stroop color-word interference test (Charchat-Fichman and Oliveira, 2009) indicate lower performance. Thus, in order to improve their graphic depiction, the direction of change was inverted (**Figure 1**). The girl's performance on the numerical-cognitive tasks was compared to that of Controls using the statistical methods for neuropsychological case studies developed by Crawford and colleagues (Crawford and Howell, 1998; Crawford and Garthwaite, 2002; Crawford et al., 2010). The analysis concerns the typicality of her performance in comparison with the Control groups. The modified *t*-test proposed by Crawford and Garthwaite (2002) calculated with singlms.exe was used to compare her scores on each task to that

of the Control groups' means. Effect size and power analyses were also calculated (Crawford et al., 2010).

RESULTS

The results are organized into four subsections: intelligence and school performance, behavioral assessment, general neuropsychological assessment and numerical-cognitive assessment.

Intelligence and School Performance

The girl performed normally on intelligence tests, reaching the 60th percentile on Raven's Colored Progressive Matrices (Angelini et al., 1999) in both assessments.

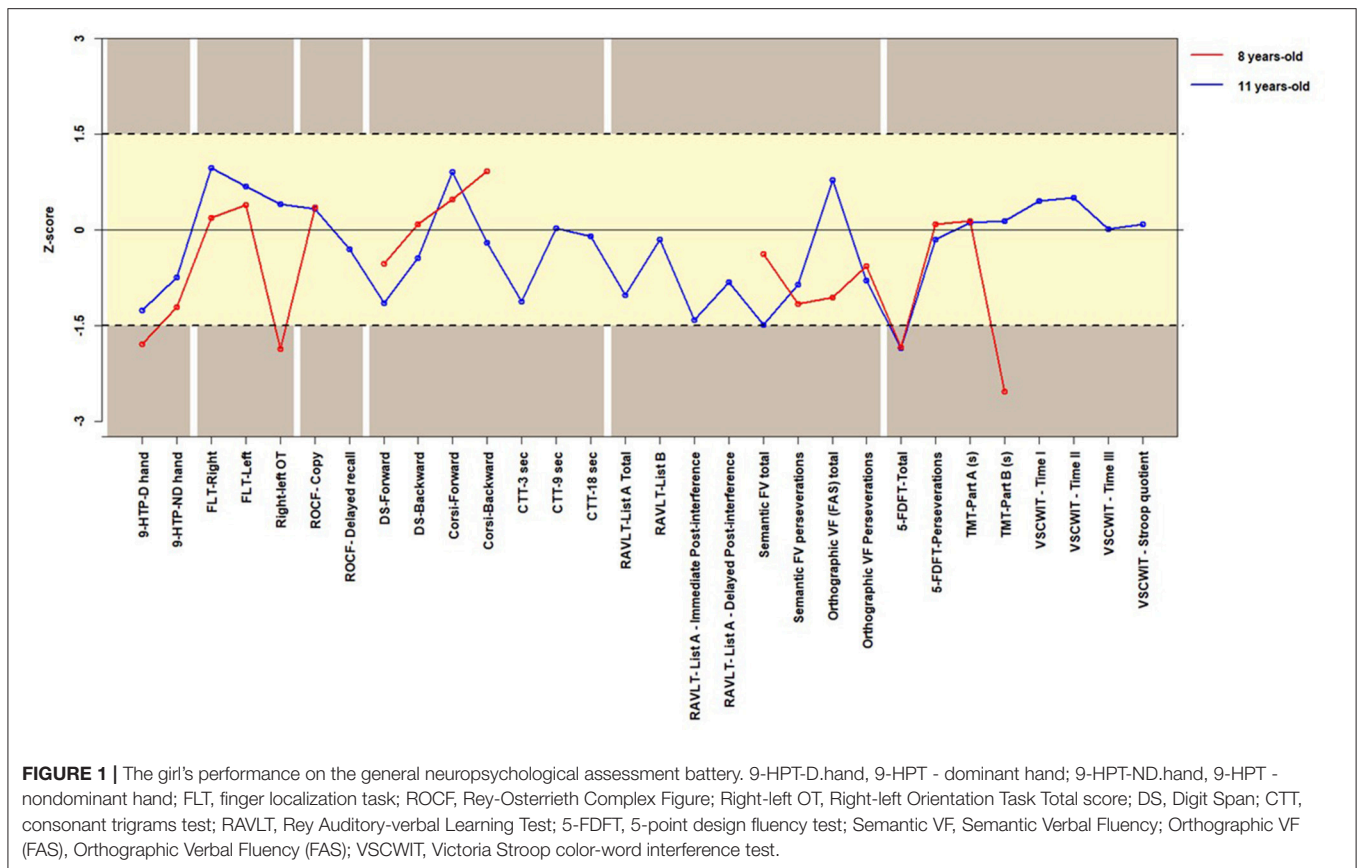
At 11 years, the girl obtained a WISC-III Full-Scale IQ of 98. Although her results did not show a discrepancy between Verbal and Performance IQ, she presented a heterogeneous profile in the subtests.

In the Verbal subscales, the girl obtained average scaled scores in the tasks that involved ability to synthesize and categorize verbal knowledge (Vocabulary: scaled score = 15; z-score = 1.66; Similarities: scaled score = 12; z-score = 0.66; and Comprehension: scaled score = 11; z-score = 0.33). Otherwise, she presented lower scaled scores, still in the normal range, in the tasks that evaluated word problem solving (Arithmetic: scaled score = 8; z-score = -0.66), general knowledge and intellectual curiosity (Information: scaled score = 6; z-score = -1.33), and verbal memory (Digit Span: scaled score = 6; z-score = -1.33). In the performance subscales, the girl obtained average scaled scores in the tasks that involved organization of the whole from separate elements (Object Assembly: scaled score = 12; z-score = 0.66), visual organization (Picture Completion: scaled score = 11; z-score = 0.33), visual memorization and motor coordination (Coding: scaled score = 11; z-score = 0.33). Additionally, she presented below average scaled scores in the tasks that evaluated capacity for visual attention (Symbol Search: scaled score = 6; z-score = -1.33), and analysis and interpretation (Picture Arrangement: scaled score = 4; z-score = -2.00).

The girl's performance on the Spelling and Reading subtests of the TDE (Stein, 1994) was classified as average (PRs between 25 and 75), at both ages of 8 and 11 years. In the TDE Arithmetic subtest, her performance was below the PR 25 at both 8 and 11 years. At 11 years, she was also evaluated with non-word repetition (Santos and Bueno, 2003), non-word reading and phoneme elision tasks (Lopes-Silva et al., 2014). The girl performed at the maximum level in these three phonological processing tasks.

Behavioral Assessment

During the evaluation, the girl was extremely shy and sometimes required extra incentive in order to participate. In the CBCL (Rocha et al., 2012), she attained clinical scores that identified social ($T = 66$), attention ($T = 73$), DSM-anxiety ($T = 65$) and DSM-ADHD problems ($T = 66$). Scores in the other subscales were in the typical range.



General Neuropsychological Assessment

At 8 years old, the girl had deficits in motor dexterity in the right (dominant) hand ($z = -1.79$) and in right-left orientation ($z = -1.87$). At 11 years old, her performance on both tasks did not differ from the performance of the Controls in motor dexterity and right-left orientation (**Figure 1**).

Visuospatial and visuoconstructional abilities were measured using the Rey-Osterrieth Complex Figure copy and delayed recall. The girl performed typically in both evaluations. At 8 years old, her performance was superior to that of the Controls (copy: $z = 0.35$). At 11 years old, her performance was similar to that of the Controls (copy: $z = 0.32$; delayed recall: $z = -0.31$) (**Figure 1**).

The girl also performed typically on short-term and working memory tasks. On the digit span, her performance was similar to that of the Controls, in both the 8-year-old (z Forward = -0.54 ; z Backward = 0.08) and 11-year-old (z Forward = -1.15 ; z Backward = -0.45) evaluations. On the Corsi Blocks, she also performed typically. Her performance was similar to the Controls in both the 8-year-old (z Forward = 0.48 ; z Backward = 0.92) and 11-year-old (z Forward = 0.90 ; z Backward = -0.21) evaluations (**Figure 1**).

At the 11-year-old evaluation, two tasks related to memory were added to the battery of neuropsychological tests. In the Consonantal Trigrams, which evaluate interference in short-term memory, and in the Rey auditory verbal learning test (RAVLT), a

task that evaluates verbal long-term memory, the girl performed similarly to the Controls.

The girl presented evidence of impairment in some executive functions in both evaluations. At 8 years old, she presented low productivity on the 5-point design fluency test, differing from the Controls ($z = -1.84$). This difference persisted in the 11-year-old assessment ($z = -1.86$). Productivity in the semantic word fluency task was typical at 8 years ($z = -0.38$) and slightly over the cut-off score at 11 years ($z = -1.49$). In part B of the Trail Making test, which evaluates motor skills, processing speed, attention capacity (visual search), monitoring, inhibition and set-shifting, she presented a much lower performance than the Controls at the 8-year-old assessment ($z = -2.54$), but no differences were found between the girl and the Controls ($z = -0.14$) at 11 years. At the 11-year-old assessment, one task was added to the battery with the purpose of evaluating the executive functions in more detail. In the Victoria Stroop color-word interference test, which evaluates monitoring, error detection/correction and inhibitory control, she presented satisfactory performance (Stroop quotient: $z = 0.08$).

Numerical-Cognitive Abilities

The results of the numerical-cognitive tasks are presented in **Table 3**. Although the SRT is not a numerical task, it was used to control for effects of general processing speed on numerical tasks. At the 8-year-old assessment, the girl's SRTs were slower

than that of the Controls ($p = 0.01$, $d = 2.33$). At the 11-year-old assessment, her performance was similar to that of the Controls ($p = 0.38$, $d = 0.30$). An efficiency score, penalizing reaction time by error rate, was used to index the results in the single-digit magnitude comparison task. No similar compensations were used for speed-accuracy trade-offs in the non-symbolic comparison (w) and set-size magnitude estimation (cv) tasks, as the emphasis on the dependent measures in these tasks is related to accuracy.

Non-symbolic magnitude comparison: In addition to her higher reaction time on the control task, the girl presented much lower performance in reaction time on the non-symbolic magnitude comparison task ($p < 0.001$, $d = 2.94$), when compared to the Controls. At 8 years, her error rate in the non-symbolic magnitude comparison task was significantly higher ($p = 0.02$, $d = 2.00$). The log-Gaussian model did not adjust at 8 years, so it was not possible to calculate the internal Weber fraction (Table 3). At 11 years old, her reaction times on the non-symbolic magnitude comparison task were slightly above the cut-off score ($p = 0.07$, $d = 1.51$), when compared to the Controls. The internal Weber fraction was 0.28 ($p = 0.06$, $d = 1.60$).

Single-digit magnitude comparison task: At 8 years, the girl presented significantly higher RTs ($p = 0.01$, $d = 2.63$) and error rates ($p = 0.03$, $d = 1.83$) in the single-digit magnitude comparison tasks, when compared to the Controls. Her efficiency score P was significantly higher than that of the Controls ($p < 0.001$, $d = 4.22$). No significant RT ($p = 0.48$, $d = -0.02$), error rate ($p = 0.22$, $d = 0.80$) or efficiency score P ($p = 0.40$, $d = -0.25$) differences were observed at 11 years in the single-digit magnitude comparison task.

Set-size estimation: At 8 years, her performance on the set-size estimation task was random. At 11 years, she presented a significantly higher coefficient of variation when compared to the Controls on the set-size estimation task ($p < 0.001$, $d = 4.75$).

Single-digit calculation: At 8 years, her performance was lower than that of the Controls on the single-digit operation tasks, both in simple addition ($p = 0.01$, $d = -2.38$) and in simple subtraction ($p = 0.01$, $d = -1.06$). At this age, the girl was unable to perform any slightly more complex addition or subtraction operations. Multiplication items were not applied at 8 years. At 11 years, her performance did not differ from the Controls in simple addition ($p = 0.07$, $d = 1.54$), complex addition ($p = 0.14$, $p = -1.12$), simple multiplication ($p = 0.33$, $d = 0.43$) and complex multiplication ($p = 0.49$, $d = -0.01$) (Table 3). Difficulties in simple subtraction ($p < 0.001$, $d = -3.92$) and complex subtraction ($p = 0.01$, $d = -2.52$) persisted.

Arabic number reading and writing: At 8 years, the girl presented much lower performance than the Controls on the Arabic number reading task ($p = 0.05$, $d = -1.66$). Her performance on the Arabic number writing task was normal ($p = 0.48$, $d = -0.05$). At 11 years, the girl's performance was adequate in tasks that assessed Arabic numbers reading ($p = 0.28$, $d = 0.58$) and writing ($p = 0.34$, $d = 0.40$).

Simple word problems: At 8 years, the girl's performance on simple word problems was below the cut-off score when compared to the Controls ($p = 0.06$, $d = -1.61$). At 11 years, her performance on this task was normal ($p = 0.18$; $d = -0.93$).

DISCUSSION

This is the first study to characterize in detail the cognitive-neuropsychological phenotype, including cognitive-numerical performance, of an individual with an atypical distal microdeletion on the long arm of chromosome 22 (22q11.2DS LCR22-4 to LCR22-5). The participant is a girl identified through a school population screening for math learning difficulties (Carvalho et al., 2014). This girl was adopted in early infancy and lived in a stable family environment. She was assessed twice, at 8 and 11 years. Her intelligence was normal average at both times. Math learning difficulties persisted from 8 to 11 years, with performance below the PR 25. No difficulties were observed in word reading, word spelling and related phonological abilities. The family reported reading comprehension difficulties. Inattention and social anxiety symptoms were also observed. General neuropsychological assessment disclosed some minor alterations. Visuospatial/visuoconstructional abilities, working memory and long-term memory were average at both times. At 8 years, she exhibited impairments in motor dexterity, right-left orientation and alertness. These impairments were not observed at the 11 years assessment. Difficulties with some executive function tasks were detected at 8 years, such as in the productivity of the 5-point-design fluency task and the set-shifting dimension of the trail-making test. These difficulties had largely disappeared by 11 years.

Persistent math learning difficulties were associated with impairments in both non-symbolic and symbolic numerical magnitude processing and in single-digit calculation. Statistically significant slower reaction times and higher error rates were observed in all non-symbolic and symbolic numerical magnitude processing tasks at 8 years. At 11 years, single-digit magnitude comparison was average, however, she exhibited difficulties with the accuracy of non-symbolic numerical representations ($d = 1.60$) and set-size estimation ($d = 4.75$). Single-digit calculation was consistently impaired at both times. At 11 years, the girl had mastered single-digit addition and multiplication calculations, but she was still struggling with even the most simple subtraction problems. She did not present difficulties with very simple word problems involving single-digit addition and subtraction, at either time. Symbolic numerical transcoding was also typically acquired.

We will discuss the main theoretical and clinical/educational issues raised by the present study in four sections: (a) neuropsychological functioning; (b) cognitive-numerical abilities; (c) mechanisms of math learning difficulties; and (d) clinical and educational implications.

Neuropsychological Functioning

Atypical 22q11.2DS (LCR22-4 a LCR22-5) is a new genetic entity, related but different from typical 22q11.2DS (LCR22-2 a LCR22-4) (Carvalho et al., 2014). Previous research consists exclusively of case (series) reports. The behavioral and cognitive profile of affected individuals was characterized only qualitatively, through clinical description. In this study, we move a step forward, reporting data from a detailed neuropsychological investigation and testing hypothesis regarding the nature of

TABLE 3 | The girl's performance on the numerical-cognitive assessment battery.

		8 years-old						11 years-old					
Measures		Girl	Controls (n = 35)		t	p	d	Girl	Controls (n = 24)		t	p	d
			Mean	SD					Mean	SD			
Simple reaction time	Reaction time (ms)	747.69	487.18	112.02	2.29	0.01	2.33	388.71	368.93	64.18	0.30	0.38	0.30
Non-symbolic magnitude comparison	Reaction time (ms)	2116.20	1281.65	284.28	2.89	<0.001	2.94	1514.72	1117.07	262.04	1.48	0.07	1.51
	Error rate	0.74	0.52	0.11	1.97	0.02	2.00	0.49	0.39	0.06	1.63	0.06	1.66
	Weber fraction (w)	–	–	–	–	–	–	0.28	0.20	0.05	1.56	0.06	1.60
Single-digit Magnitude Comparison Task	Reaction time (ms)	1706.94	1041.76	235.50	2.59	0.01	2.63	770.62	776.24	197.92	–0.02	0.48	–0.02
	Error rate	0.57	0.35	0.12	1.80	0.03	1.83	0.24	0.20	0.05	0.78	0.22	0.80
	Efficiency score P	2715.59	1290.35	337.61	4.16	<0.001	4.22	858.19	928.87	280.70	–0.24	0.40	–0.25
Set-size Magnitude Estimation	Coefficient of variation (cv)	–	–	–	–	–	–	0.33	0.14	0.04	4.65	<0.001	4.75
Arabic number reading task	Accuracy	20	25.53	3.34	–1.63	0.05	–1.66	28	27.67	0.56	0.57	0.28	0.58
Arabic number writing task	% correct	78.57	79.62	21.32	–0.05	0.48	–0.05	100	99.55	1.11	0.39	0.34	0.4
Single-digit operations	Simple addition	2	9.69	3.23	2.34	0.01	–2.38	11	11.71	0.46	–1.51	0.07	1.54
	Complex addition	0	7.22	4.02	1.77	0.04	–1.80	10	12.71	2.40	–1.1	0.14	–1.12
	Simple subtraction	1	8.53	3.34	2.22	0.01	–2.25	5	11.04	1.54	–3.84	<0.001	–3.92
	Complex subtraction	0	4.69	4.43	1.04	0.15	–1.06	0	8.58	3.40	–2.47	0.01	–2.52
	Simple multiplication	–	–	–	–	–	–	15	13.92	2.51	0.42	0.33	0.43
	Complex multiplication	–	–	–	–	–	–	6	6.04	4.00	–0.01	0.49	–0.01
	Simple word problems	4	8.81	2.99	1.50	0.06	–1.61	9	10.46	1.56	–0.91	0.18	–0.93

Bold value indicates Statistical significance: $p < 0.05$.

observed cognitive-numerical impairments. We first discuss the results of the general neuropsychological assessment.

Intelligence: In general, most cases of 22q11.2DS (LCR22-4 to LCR22-5) have been described as having intellectual disability and receiving special education (Ben-Shachar et al., 2008; Xu et al., 2008; Mikhail et al., 2014; Lindgren et al., 2015). Only one study reported the IQ of a girl with microdeletion in LCR22-4 to LCR22-5 region. In this study, Verhoeven et al. (2011) described a 17-year-old female and her level of intelligence was found to be borderline (total WISC-R IQ=73). Two cases presenting presumably normal intelligence without detailed description were reported by Ben-Shachar et al. (2008) and Fagerberg et al. (2012).

In children with typical 22q11.2DS, intellectual disability is present in 40% to 45% of affected individuals. When intelligence is normal, usually the IQ is in the borderline range (IQ = 70 to 85, Swillen et al., 1997; Woodin et al., 2001; Green et al., 2009). In children, lower scores are observed in the Performance IQ. This discrepancy tends to decrease in adults (Moberg et al., 2018). One hypothesis is that concomitant lowering of Verbal IQ tends to reduce the discrepancy. A reduction of Verbal IQ from childhood to adolescence has been reported in some individuals with typical 22q11.2DS, and it is considered a risk factor for psychosis (Gothelf et al., 2005, 2009).

General intelligence scores remained stable in this girl for three years. Further follow-up is required. Normal intelligence

in our participant indicates that intellectual disability is not an necessary phenotypic trait in 22q11.2 (LCR22-4 to LCR22-5). Research on intellectual abilities of individuals with genetic syndromes is biased by the fact that most severe cases have a higher probability of being recognized by families, clinicians and educators.

Visuospatial and motor abilities: Previous reports have underscored the severity of impairments in motor dexterity and visuospatial/visuoconstructional processing in cases of 22q11.2DS (LCR22-4 to LCR22-5). Lindgren et al. (2015) described a 4-year-old patient with 22q11.2DS (LCR22-4 to LCR22-5), that presented deficits in visual perception and motor integration, and mildly delayed gross motor milestones. In 2008, Rodningen and coworkers briefly described a 7-year-old patient with 22q11.2 (LCR22-4 to LCR22-5), presenting the same profile. Additional cases of 22q11.2DS (LCR22-4 to LCR22-5) showing motor deficits have been reported in the literature (Ben-Shachar et al., 2008; Beaujard et al., 2009; Verhoeven et al., 2011; Fagerberg et al., 2012; Mikhail et al., 2014; Spineli-Silva et al., 2017). Impairments in visuomotor integration were reported in two additional articles (Mikhail et al., 2007; Verhoeven et al., 2011).

Individuals with typical 22q11.2DS also present motor delays and difficulties with motor coordination from infancy on (Swillen et al., 1999; Bearden et al., 2001; Gerdes et al., 2001; Vicari et al., 2011). Large and consistent deficits were found for motor skills ($d = -1.17$) (Moberg et al., 2018). Additionally, occurrence of visuospatial and visuoconstructional impairments is frequent although variable in typical 22q11.2DS (Antshel et al., 2008; Jacobson et al., 2010; Schoch et al., 2014).

Most individuals previously reported with atypical 22q11.2DS were observed in infancy and at preschool age. Unfortunately, as our participant was adopted, there is no information regarding her obstetric and early infancy developmental background. The family reports motor delay at the end of the first year, when she was adopted. This improved in the following 3 months. Minor impairments in motor dexterity, body representation and alertness were observed at 8 years and improved with time (**Figure 1**). Additionally, and importantly, she did not present visuospatial/visuoconstructional impairments at either time (**Figure 1**). Anyway, the severity of visuospatial and motor impairments in previous reports of both atypical and typical 22q11.2DS contrast with the mildness of impairments in our participant.

Memory: Memory functions were not investigated in previous reports of atypical 22q11.2DS.

In general, individuals with typical 22q11.2DS present better performance on tasks of verbal rather than visuospatial memory (Woodin et al., 2001; Wong et al., 2014). However, both kinds of memory are impaired compared to controls. Moderate to large effect sizes were found for verbal memory ($d = -0.70$) and visual memory ($d = -1.0$) (Moberg et al., 2018). Individuals with typical 22q11.2DS present similar performance as controls in tasks of information acquisition and retrieval (Lajiness-O'Neill et al., 2006; Debbané et al., 2008). Difficulties are more apparent in tasks in which the participant needs to discriminate stimulus relevance. These memory alterations may constitute a trait

vulnerability marker signaling increased risk for schizophrenia in the typical 22q11.2DS population (Debbané et al., 2008).

Working and episodic visuospatial memories were intact in our participant (**Figure 1**). A discrepancy between higher digit span scores and lower but still normal total WISC Digit scores was observed and may be ascribed to attentional fluctuation. Difficulties with attention were also qualitatively observed in the RAVLT performance.

Executive functions: Deficits in executive functions were described in the case of 22q11.2DS (LCR22-4 to LCR22-5) reported by Verhoeven et al. (2011). They described an 18-year-old girl with borderline intelligence and deficits related to planning and concentration. Other reported cases have presented more severe cognitive impairments related to intellectual disability.

Impairments in executive functions are frequent, severe and persistent in individuals with typical 22q11.2DS (Woodin et al., 2001; Robin and Shprintzen, 2005). Moberg et al. (2018) observed moderate to large impairments in basic executive functions (up to $d = -0.90$). Executive function impairments, together with progressive verbal IQ decline, may play a role in the vulnerability to psychiatric disorders, such as psychoses (Gothelf et al., 2005, 2009).

The girl presented difficulties with some executive function tasks. We feel that her deficits in executive functions were slight and tended to improve. In the 3-year period of observation, no deterioration in her cognitive status was observed.

Psychosocial functioning: The girl is the eighth case with 22q11.2DS (LCR22-4 to LCR22-5) reported in the literature presenting symptoms of impulsivity and inattentiveness (Mikhail et al., 2007; Fagerberg et al., 2012). Her psychosocial functioning profile, including attention, social and anxiety problems, had some similarities and differences with those reported previously. Mikhail et al. (2014) described 4 cases with social immaturity, poor impulse control and anger issues, ADHD, anxiety and Asperger Disorder. The girl did not present symptoms of autism, but she presented characteristics of social phobia. Aggressive behaviors also seem to be common in patients with 22q11.2DS (LCR22-4 to LCR22-5) (Ben-Shachar et al., 2008; Verhoeven et al., 2011; Mikhail et al., 2014; Lindgren et al., 2015). Aggressive behavior was not a major issue in the participant.

The relatively mild psychosocial impairment observed in our participant contrasts with the more severe difficulties encountered by individuals with both atypical and typical 22q11.2DS, including the risk of psychosis (Bassett and Chow, 2008). In typical 22q11.2DS, psychosis is estimated to occur in up to 22.6% of patients after adolescence (Bassett and Chow, 2008).

School learning difficulties: Normal intelligence and math learning difficulties have been described in two cases of 22q11.2DS (LCR22-4 to LCR22-5) (Verhoeven et al., 2011; Carvalho et al., 2014). The most salient phenotypic features presented by this participant were the difficulties with number processing and arithmetic calculation. This is the first study to report a detailed neuropsychological investigation of an individual with 22q11.2DS (LCR22-4 to LCR22-5) with normal intelligence and specific learning difficulties.

In summary, the present study suggests a huge variability in the cognitive and behavioral phenotype of 22q11.2DS (LCR22-4 to LCR22-5). Less severely affected individuals may have normal intelligence associated with milder behavioral issues and specific school learning problems. Next, we compare these math learning difficulties with those observed in typical 22q11.2DS (LCR22-2 to LCR22-4). Math learning difficulties will be emphasized, as they are a prominent feature of the present participant as well as in typical 22q11.2DS.

Cognitive-Numerical Abilities

It is interesting to compare the profile of cognitive-numerical and arithmetic performance observed in the girl with that of typical 22q11.2DS. Math learning difficulties are a hallmark of the 22q11.2DS phenotype in individuals with normal intelligence (De Smedt et al., 2009). Math learning difficulties in typical 22q11.2DS seem to be unrelated to phonological processing impairments and probably reflect difficulties in more basic numerical and/or visuospatial processing (De Smedt et al., 2008).

De Smedt et al. (2009) observed that 22q11.2DS children's performance did not differ from that of controls in the tasks of reading numbers and single digit calculation. However, 22q11.2DS children were slower than controls in number comparison and in addition/subtraction calculations with larger numbers.

Oliveira et al. (2014) were the first to report inaccuracy of non-symbolic numerical magnitude representations (indexed by the internal Weber fraction, w) in typical 22q11.2DS. However, performance was variable, as not all individuals with 22q11.2DS presented impairments in ANS accuracy. Impairment in ANS, indexed by w in the non-symbolic numerical comparison task, was later confirmed by Attout et al. (2017). Additionally, these authors observed that ANS accuracy was impaired in the visuospatial but not in the auditory version of the non-symbolic comparison task. This suggests a connection between non-symbolic numerical and visuospatial representations. As mentioned before, visuospatial impairments are an important feature of typical 22q11.2DS.

A connection between numerical and spatial representations is suggested by the mental number line model of approximate numerical representations (Dehaene, 1997, 2007; Nieder and Dehaene, 2009). According to this model, the psychophysical signature of numerical magnitude representations suggests a spatialization of approximate numerical representations: (a) numerical magnitude discriminations are increasingly (ratio variability) and proportionally (scalar variability) more difficult as the distance between the numerical stimuli decreases; (b) accuracy in numerical representations also decreases as the numerical magnitude increases in a logarithmically compressed way; finally, (c) smaller digits are processed preferentially by the right and larger digits by the left hemispheres, suggesting a spatial orientation of the mental number line. According to Dehaene (2007) and Nieder and Dehaene (2009), these characteristics indicate that non-symbolic numbers may be represented approximately as a log-Gaussian distribution of the neuronal discharges ordered by numerical magnitudes.

The spatial nature of numerical representations and their impairments in typical 22q11.2DS have been explored in several studies by Simon et al. (2005a,b) and Simon (2008). In these studies, impaired performance of children with 22q11.2DS in a non-symbolic comparison task was associated with visuospatial manipulations reducing stimuli discriminability. According to the granularity hypothesis, Simon (2008) attributed the numerical processing deficits of individuals with 22q11.2DS to a more basic spatial representation inaccuracy or lack of spatial resolution.

Our participant presented persistent math difficulties, investigated from 8 to 11 years. Four possible cognitive-numerical sources for these difficulties may be considered: (a) visuospatial and visuoconstructional impairments; (b) phonological processing impairment; (c) basic numerical impairment; (d) executive dysfunction. The first two are discarded because there was no evidence of impairment in visuospatial/visuoconstructional and phonological processing abilities. Transcoding abilities of more complex numerals is indicative of good spatial and phonological processing abilities. Moreover, improving ability with commutative single-digit operations and persisting difficulties with subtraction suggest an impairment in the ANS. This hypothesis will be considered next.

In the present participant, the agreement among impairments of numerical processing in different modalities and tasks and their persistence is remarkable. Some evidence indicates that experimental tasks of numerical processing lack concurrent validity (Maloney et al., 2010; Price et al., 2012; Pinheiro-Chagas et al., 2014; Smets et al., 2015) and their test-retest reliability has not been explored extensively (Haase et al., 2014). The results indicate that, at least in some cases, basic numerical impairments may be consistent and persistent.

The most remarkable feature of numerical-cognitive impairments in the girl is related to severe impairments in basic numerical magnitude processing. The available data do not allow us to definitely decide if her impairments are related to non-symbolic numerical magnitude representational inaccuracy (Landerl et al., 2004) or to access to non-symbolic representations from symbolic ones (Rousselle and Noël, 2007). Accordingly, an individual could have difficulties learning math owing to some basic numerical magnitude representational deficit or to difficulties with accessing, storing and manipulating numerical information in working memory. These hypotheses will be addressed in the next section, in the context of the mechanisms putatively involved in MD.

Cognitive Mechanisms of Math Learning Difficulties

No substantive qualitative differences were observed in the cognitive mechanisms putatively underlying the present participant's math difficulties and those observed multifactorial developmental math learning difficulties (Wilson and Dehaene, 2007; Karagiannakis et al., 2014). The mathematical behavioral genetic approach partitions variance at the population level and does not allow identification of specific mechanisms implicated in single individuals. This can be accomplished only

by molecular-genetic and neuropsychological investigations of specific genetic etiologies.

Current multiple deficit models of developmental disabilities consider that the phenotypic expression is dependent on complex genetic-environmental interactive mechanisms (Pennington, 2006; Johnson, 2012). Relationships between the genetic-environmental etiologic level and the phenotypic expression are not simple, one-to-one, and are subject to environmental sources of regulation at different times. The construct endophenotype was suggested to characterize intermediate steps in this complex, epigenetic path from the genotype to the phenotype (Rutter et al., 2006; Bishop, 2009).

Several endophenotypes were identified in the present study as potentially relevant for the girl's math difficulties as well as for math difficulties in general. In addition to basic numerical processing, discussed in the last section, the following potentially relevant mechanisms were identified in the present participant:

Motor ability: Basic perceptual and motor impairments are a frequent observation in several developmental disorders (Denckla, 1997, 2003), and are predictive of cognitive and behavioral problems at school age (Batstra et al., 2003). Deficits in finger gnosis (Costa et al., 2011) and motor incoordination (Lonnemann et al., 2011) have been described in children with MD. The meaning of these perceptual and motor impairments is uncertain. Bottom-up theories interpret cognitive deficits as a consequence of a disordered developmental process, encompassing the most basic perceptual motor abilities from infancy on (Nicolson and Fawcett, 2010; Elliott and Grigorenko, 2014). According to the procedural deficit hypothesis, MD could be related to difficulties in automatizing the implicit associations underlying numerical concepts and operations (Vandervort, 2017; Prado, 2018). An alternative explanation is that perceptual and motor impairments constitute markers of severity or co-localizes, indicating the presence and anatomic location of brain dysfunction (Denckla, 1997, 2003).

Working memory and executive functions: Impairments in working memory (Raghubar et al., 2010) and executive functions (Bull and Lee, 2014) are an important trait identified in individuals with MD. The ability to store and manipulate information temporarily in working memory is an important requirement at every step in the acquisition of arithmetics, such as counting (Geary et al., 2004), single-digit calculation (Menon et al., 2000; De Smedt et al., 2009), multi-digit calculation (Klein et al., 2009), numerical transcoding, (Barrouillet et al., 2004; Camos, 2008) and word problem solving (Swanson and Sachse-Lee, 2001). Attentional and executive functions have been implied, even in basic quantitative-numerical decisions (Clayton and Gilmore, 2015; Merkley et al., 2016). For example, inhibition of irrelevant perceptual dimensions may play a role in non-symbolic numerical magnitude comparisons. It is notoriously difficult to experimentally control covariation between the discrete numerical and continuous dimensions of stimuli in these tasks (Leibovich and Henik, 2014). The difficulty of the task could then be related to the need to inhibit the irrelevant continuous dimensions, such as surface and luminance, in order to decide based on the relevant discrete magnitude dimension. Other research indicates, however, that in the range of numerosities

usually investigated, discrete numerosity is more perceptually salient and associated with math achievement than continuous dimensions such as texture (Anobile et al., 2016).

Math anxiety: Math anxiety is weakly and negatively associated with math achievement, with correlations on the order of -0.25 to -0.40 (Hembree, 1990). Math anxiety is both a risk factor and a consequence of MD (Ma, 1999). However, math anxiety and achievement are dissociable phenomena (Lee, 2009; Stankov et al., 2012), with both highperforming individuals being anxious and lowperforming individuals not being anxious. Usually, math anxiety is not considered a sort of learning disability (Ashcraft and Krause, 2007). It is considered an important concomitant or aggravating factor of existing difficulties.

In summary, several mechanisms were identified as potentially relevant for the MD in the present participant. It is important to balance and to integrate the evidence, connecting it to the big picture of MD in general. It is unfortunate that the genetic and psychosocial background of the participant before adoption is unknown. Data indicates that adopted children have been previously subject to both genetic and environmental risks for poor school achievement (Van Ijzendoorn et al., 2005).

Perceptual and motor impairments and anxiety may also have played a role in the genesis of the girl's math difficulties. Right-left orientation difficulties and motor dexterity improved with time but could have played a role at a crucial moment in learning arithmetics. Math anxiety may have competed for cognitive resources required for math learning at several moments.

The most interesting question is the relative role played by basic numerical processing and executive functioning. The possibility that executive dysfunction may have played a role cannot be excluded. First, her difficulties with executive functions were relatively mild, at least at the times of assessment; and, the clinical history does not suggest severe impairments in self-regulation. Second, her basic numerical processing difficulties were severe, persistent and concordant across modalities and tasks.

The numerical processing abilities of the participant can be interpreted in terms of the criteria proposed by Rousselle and Noël (2007). According to these authors, an access disorder, probably related to executive dysfunction, is characterized by variable and discrepant performance, with sparing of non-symbolic over symbolic numerical processing. The representational deficit is otherwise characterized by modality-independent and comprehensive difficulties with numerical processing. The pervasiveness of the girl's numerical processing difficulties and the mildness of her executive function difficulties suggest a representational deficit.

Investigations at the population and single individual level play complementary roles in partitioning variance and identifying specific sources of difficulties in math achievement. Since working memory and executive function impairments are frequent in all developmental disorders, one important question is related to the specificity of the problem. Why should one kid develop difficulties only in math and the other only in reading?

Multiple deficit models help to understand the complex interplay between specific and general cognitive factors

in the origin of MD. According to a model proposed by Johnson (2012), a kid with a basic numerical processing impairment could compensate for the resulting difficulties, if executive processing resources are available. Otherwise, when general processing resources are insufficient, the difficulties are not compensated and may call attention of parents, educators and clinicians, leading to a diagnosis. In the present participant, multiple sources of cognitive and psychosocial variability were identified that could interact with the genetic condition, leading to math learning difficulties.

Clinical and Educational Implications

The main results of our study are that math learning difficulties may be associated with a specific genetic etiology (22q11.2DS; LCR22-4 to LCR22-5) and with more or less specific cognitive mechanisms (ANS and/or executive function impairments). Obviously, identifying a potential specific genetic etiology in a case of MD does not ensure that it plays a causal role in the difficulties of that single individual. It also does not exclude a role for other genetic or environmental factors. It is especially important to consider this in the present individual, as the girl was adopted and little information is available on her background before adoption. What are the implications of these findings for neuropsychological and educational practice?

Etiology of developmental and learning disorders is considered to be multifactorial; i.e., resulting from the interaction of several polygenic and environmental influences (Asbury and Plomin, 2013). It is, however, increasingly being recognized that, at the individual level, specific causes may play a role (Carvalho et al., 2014). For example, chromosomal aneuploidies have been recognized as a cause of language development and reading learning difficulties (Simpson et al., 2014). Specific genetic causes also contribute to autism (Cohen et al., 2005). The extreme variability of clinical presentation makes diagnosis difficult in milder cases.

Other research indicates that individuals with learning difficulties present higher rates of medical, especially neurological and psychiatric, comorbidities. This may occur in math (Shalev and Gross-Tsur, 1993) although not in reading learning difficulties (Cuvellier et al., 2004; Billard et al., 2008). Focal cerebral damage has been reported in cases of developmental dyscalculia and dyslexia (Levin et al., 1996; Daigneault and Braun, 2002). Rolandic epilepsy is commonly associated with learning difficulties in children of normal intelligence (Canavese et al., 2007). Common diseases, such as diabetes and asthma, are also more common in children with learning difficulties than in the general population (Blackman and Gurka, 2007; Hannonen et al., 2010).

Specific etiologies might be more common than usually thought. They are not identified because they are not looked for. Polygenes play a causative role at the population but not at the single individual level. The same holds for psychosocial factors. Deprivation, neglect or maltreatment are the most important risk factors for psychopathology and learning difficulties at the population level (Altarac and Saroha, 2007; Belsky, 2007). In a single individual, it is often difficult to establish a causative role

for these psychosocial influences, as not all individuals subject to a risk present the outcome (Caspi et al., 2003; Nobile et al., 2010).

Even if the occurrence of a specific etiology were an infrequent event, underdiagnosis has important consequences, as the individual is deprived of proper health and educational counseling. This is especially important in the era of response to intervention (RTI). Learning difficulties are increasingly being handled by teachers in the schools, using the RTI approach, without referral to specialists (Hale et al., 2010). In the RTI approach, it may take several semesters until teachers recognize that a kid presents more severe and stable difficulties that do not respond to the interventions. Furthermore, they may be associated with a higher probability of a genetic etiology. Referrals for specialized diagnosis and care may be delayed for these individuals.

We argue that teachers must be aware of the possibility that children with learning difficulties are a group at risk for several medical, neurological and psychiatric conditions. Our results suggest that math learning difficulties may function as a kind of red-flag, pointing to possible genetic etiologies. Some red-flags for genetic syndromes may be minor, albeit observable by teachers: short or tall stature, congenital malformations, hypotonia, poor motor coordination, anomalous handedness, history of developmental delay, etc. “Funny face” is an important red-flag. These children have no facial malformations but, rather, small, subtle dysmorphisms such as a low nasal bridge, markedly upslanting or downslanting palpebral fissures, small or prominent chin, low set ears, etc. (Huang et al., 2010). Normal people may have one or two such dysmorphisms, but they are not enough to characterize a “funny face.” Minor motor impairments may also hint at a neurological etiology (Daigneault and Braun, 2002; Batstra et al., 2003). Adoption is another important risk factor for developmental disorders of genetic or environmental etiology (Altarac and Saroha, 2007; Tenenbaum et al., 2011). However, it is important not to forget that most children with math learning difficulties will have a perfectly normal constitution and no genetic syndrome.

Finally, our research design has no power to establish a definite role for ANS over executive function impairments in the etiology of the girl's math learning difficulties. Results indicate however, that specific mechanisms, such as ANS and/or executive function impairments vs. phonological and/or visuospatial/visuoconstructional processing, may play a role in specific individuals.

Again, in a given individual, it may difficult to reliably identify which cognitive mechanisms underlie the difficulties. Our own experience has been that, in accordance with the multiple deficits hypothesis, specific and general cognitive impairments interact in complex ways (Haase et al., 2014; Júlio-Costa et al., 2015; Gomides et al., 2018). Identification of the putative mechanisms is relevant for the planning of more efficient interventions (Gomides et al., 2018). Anyway, alone or interacting with general cognitive impairments, ANS may play a role in math learning difficulties. Future research should address the specific mechanisms and crucial developmental period(s) of the ANS involvement with math learning, as well as intervention strategies.

This investigation of a girl with 22q11.2DS (LCR22-4 to LCR22-5), allows us to raise the following points: (a) specific genetic alterations, such as atypical 22q11.2DS, may be related to math learning difficulties in individuals with normal intelligence and slight phenotypic traits that would remain otherwise unrecognized; (b) math learning difficulties may be severe and persistent in these cases, involving both non-symbolic and symbolic numerical magnitude processing, and eventually be associated with executive dysfunctions; (c) although the microdeleted regions in typical and atypical cases of 22q11.2 are non-overlapping, their phenotypic traits may be broadly shared, suggesting long-range interactions and complexity of genotype-phenotype associations (Zeitz et al., 2013); (d) numerical-cognitive impairments were dissociated from spared visuospatial abilities, suggesting heterogeneity of neurogenetic underpinnings. Further studies have the challenge of showing more evidence for these issues.

AUTHOR CONTRIBUTIONS

VH and MC delineated the study; LO, AJ-C, and VH conducted the neuropsychological evaluation; FS and MC conducted the genetic analyses. All authors contributed in analysing the results

and writing the paper. All authors read the final version of the paper and agree with the content of the manuscript.

FUNDING

This study was supported by grants from the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG, APQ-02755-SHA, APQ-03289-10, APQ-02953-14, APQ-03642-12). VH is supported by a CNPq fellowship (409624/2006-3, 308157/2011-7, 308267/2014-1) and Programa de Capacitação em Neuropsicologia do Desenvolvimento (FEAPAEs-MG, APAE-BH, PRONAS-Ministério da Saúde, Brasil). MC is supported by a CNPq fellowship (312068/2015-8). LO, AJ-C, and FS are supported by PhD fellowships from the Coordenação de Aperfeiçoamento de Pessoal de Ensino Superior (CAPES).

ACKNOWLEDGMENTS

The authors thank to the children, their parents, and also to the principals of the schools for taking part in this research. We thank to Mr. Peter Laspina, from ViaMundi Idiomas e Traduções for reviewing this manuscript.

REFERENCES

- Altarc, M., and Saroha, E. (2007). Lifetime prevalence of learning disability among US children. *Pediatrics* 119(Suppl. 1), S77–S83. doi: 10.1542/peds.2006-2089L
- Angelini, A. L., Alves, I. C. B., Custódio, E. M., Duarte, W. F., and Duarte, J. L. M. (1999). *Matrizes Progressivas Coloridas de Raven: Escala Especial*. Manual. São Paulo: CETEPP.
- Anobile, G., Castaldi, E., Turi, M., Tinelli, F., and Burr, D. C. (2016). Numerosity but not texture-density discrimination correlates with math ability in children. *Dev. Psychol.* 52:1206. doi: 10.1037/dev0000155
- Antshel, K. M., Fremont, W., and Kates, W. R. (2008). The neurocognitive phenotype in velo-cardio-facial syndrome: a developmental perspective. *Dev. Disabil. Res. Rev.* 14, 43–51. doi: 10.1002/ddrr.7
- Asbury, K., and Plomin, R. (2013). *G is for Genes: The Impact of Genetics on Education and Achievement*. Chichester, UK: Wiley Blackwell. doi: 10.1002/9781118482766
- Ashcraft, M. H., and Krause, J. A. (2007). Working memory, math performance, and math anxiety. *Psychon. Bull. Rev.* 14, 243–248. doi: 10.3758/BF03194059
- Attout, L., Noël, M. P., Vossius, L., and Rousselle, L. (2017). Evidence of the impact of visuo-spatial processing on magnitude representation in 22q11.2 microdeletion syndrome. *Neuropsychologia* 99, 296–305. doi: 10.1016/j.neuropsychologia.2017.03.023
- Barnea-Goraly, N., Eliez, S., Menon, V., Bammer, R., and Reiss, A. L. (2005). Arithmetic ability and parietal alterations: a diffusion tensor imaging study in velocardiofacial syndrome. *Cogn. Brain Res.* 25, 735–740. doi: 10.1016/j.cogbrainres.2005.09.013
- Barrouillet, P., Camos, V., Perruchet, P., and Seron, X. (2004). ADAPT: A developmental, asemantic, and procedural model for transcoding from verbal to Arabic numerals. *Psychol. Rev.* 111, 368–394. doi: 10.1037/0033-295X.111.2.368
- Bassett, A. S., and Chow, E. W. (2008). Schizophrenia and 22q11.2 deletion syndrome. *Curr. Psychiatry Re.* 10:148. doi: 10.1007/s11920-008-0026-1
- Batstra, L., Neeleman, J., and Hadders-Algra, M. (2003). The neurology of learning and behavioural problems in pre-adolescent children. *Acta Psychiatr. Scand.* 108, 92–100. doi: 10.1034/j.1600-0447.2003.00127.x
- Bearden, C. E., Woodin, M. F., Wang, P. P., Moss, E., McDonald-McGinn, D., Zackai, E., et al. (2001). The neurocognitive phenotype of the 22q11.2 deletion syndrome: selective deficit in visual-spatial memory. *J. Clin. Exp. Neuropsychol.* 23, 447–464. doi: 10.1076/jcen.23.4.447.1228
- Beaujard, M. P., Chantot, S., Dubois, M., Keren, B., Carpentier, W., Mabboux, P., et al. (2009). Atypical deletion of 22q11.2: detection using the FISH TBX1 probe and molecular characterization with high-density SNP arrays. *Eur. J. Med. Genet.* Year 52, 321–327. doi: 10.1016/j.ejmg.2009.05.010
- Belsky, J. (2007). Experience in childhood and the development of reproductive strategies. *Acta Psychol. Sin.* 39, 454–468.
- Ben-Shachar, S., Ou, Z., Shaw, C. A., Belmont, J. W., Patel, M. S., Hummel, M., Amato, S., et al. (2008). 22q11.2 distal deletion: a recurrent genomic disorder distinct from DiGeorge syndrome and velocardiofacial syndrome. *Am. J. Hum. Genet.* 82, 214–221. doi: 10.1016/j.ajhg.2007.09.014
- Billard, C., Fluss, J., Ducot, B., Warszawski, J., Ecalle, J., Magnan, A., et al. (2008). Study of causal factors of reading impairment in a sample of 1062 7 to 8-year-old children. *Arch. Pédiatrie* 15, 1058–1067. doi: 10.1016/j.arcped.2008.02.020
- Bishop, D. V. (2009). Genes, cognition, and communication. *Ann. N.Y. Acad. Sci.* 1156, 1–18. doi: 10.1111/j.1749-6632.2009.04419.x
- Blackman, J. A., and Gurka, M. J. (2007). Developmental and behavioral comorbidities of asthma in children. *J. Dev. Behav. Pediatr.* 28, 92–99. doi: 10.1097/01.DBP.0000267557.80834.e5
- Brankaer, C., Ghesquière, P., De Wel, A., Swillen, A., and De Smedt, B. (2017). Numerical magnitude processing impairments in genetic syndromes: a cross-syndrome comparison of Turner and 22q11.2 deletion syndromes. *Dev. Sci.* 20:e12458. doi: 10.1111/desc.12458
- Bruandet, M., Molko, N., Cohen, L., and Dehaene, S. (2004). A cognitive characterization of dyscalculia in Turner syndrome. *Neuropsychologia* 42, 288–298. doi: 10.1016/j.neuropsychologia.2003.08.007
- Bruce, S., Hannula-Jouppi, K., Puoskari, M., Franssón, I., Simola, K. O., Lipsanen-Nyman, M., et al. (2010). Submicroscopic genomic alterations in Silver-Russell syndrome and Silver-Russell-like patients. *J. Med. Genet.* 47, 816–822. doi: 10.1136/jmg.2009.069427
- Bull, R., and Lee, K. (2014). Executive functioning and mathematics achievement. *Child Dev. Perspect.* 8, 36–41. doi: 10.1111/cdep.12059
- Burnside, R. D. (2015). 22q11.21 deletion syndromes: a review of proximal, central, and distal deletions and their associated features. *Cytogenet. Genome Res.* 146, 89–99. doi: 10.1159/000438708

- Camos, V. (2008). Low working memory capacity impedes both efficiency and learning of number transcoding in children. *J. Exp. Child Psychol.* 99, 37–57. doi: 10.1016/j.jecp.2007.06.006
- Canavese, C., Rigardetto, R., Viano, V., Vittorini, R., Bassi, B., Pieri, I., et al. (2007). Are dyslexia and dyscalculia associated with Rolandic epilepsy? A short report on ten Italian patients. *Epileptic Disord.* 9, 432–436. doi: 10.1684/epd.2007.0138
- Carvalho, M. R. S., Vianna, G., Oliveira, L. F. S., Aguiar, M. J. B., Zen, P., and Haase, V. G. (2014). Are 22q11.2 distal deletions associated with math difficulties? *Am. J. Med. Genet. A* 164, 2256–2262. doi: 10.1002/ajmg.a.36649
- Caspi, A., Sugden, K., Moffitt, T. E., Taylor, A., Craig, I. W., Harrington, H., et al. (2003). Influence of life stress on depression: moderation by a polymorphism in the 5-HTT gene. *Science* 301, 386–389. doi: 10.1126/science.1083968
- Charchat-Fichman, H., and Oliveira, R. M. (2009). Performance of 119 Brazilian children on Stroop paradigm: victoria version. *Arq. Neuropsiquiatr.* 67, 445–449. doi: 10.1590/S0004-282X2009000300014
- Chen, Q., and Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychol.* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Clayton, S., and Gilmore, C. (2015). Inhibition in dot comparison tasks. *ZDM* 47, 759–770. doi: 10.1007/s11858-014-0655-2
- Cohen, D., Pichard, N., Tordjman, S., Baumann, C., Burglen, L., Excoffier, E., et al. (2005). Specific genetic disorders and autism: clinical contribution towards their identification. *J. Autism Dev. Disord.* 35, 103–116. doi: 10.1007/s10803-004-1038-2
- Costa, A. J., Lopes-Silva, J. G., Pinheiro-Chagas, P., Krininger, H., Lonnemann, J., Willmes, K., et al. (2011). A hand full of numbers: a role for offloading in arithmetics learning. *Front. Psychol.* 2:368. doi: 10.3389/fpsyg.2011.00368
- Crawford, J. R., and Garthwaite, P. H. (2002). Investigation of the single case in neuropsychology: Confidence limits on the abnormality of test scores and test score differences. *Neuropsychologia* 40, 1196–1208. doi: 10.1016/S0028-3932(01)00224-X
- Crawford, J. R., Garthwaite, P. H., and Porter, S. (2010). Point and interval estimates of effect sizes for the case controls design in neuropsychology: Rationale, methods, implementations, and proposed reporting standards. *Cogn. Neuropsychol.* 27, 245–260. doi: 10.1080/02643294.2010.513967
- Crawford, J. R., and Howell, D. C. (1998). Comparing an individual's test score against norms derived from small samples. *Clin. Neuropsychol.* 12, 482–486.
- Cuvellier, J. C., Pandit, F., Casalis, S., Lemaître, M. P., Cuisset, J. M., Platot, A., et al. (2004). Analyse d'une population de 100 enfants adressés pour troubles d'apprentissage scolaire. *Archives de Pédiatrie* 11, 201–206. doi: 10.1016/j.arcped.2003.12.004
- Daigneault, S., and Braun, C. M. (2002). Pure severe dyslexia after a perinatal focal lesion: evidence of a specific module for acquisition of reading. *J. Dev. & Behav. Pediatr.* 23, 256–265. doi: 10.1097/00004703-200208000-00011
- De Smedt, B., and Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *J. Exp. Child Psychol.* 108, 278–292. doi: 10.1016/j.jecp.2010.09.003
- De Smedt, B., Noël, M. P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends Neurosci. Educ.* 2, 48–55. doi: 10.1016/j.tine.2013.06.001
- De Smedt, B., Reynvoet, B., Swillen, A., Verschaffel, L., Boets, B., and Ghesquière, P. (2009). Basic number processing and difficulties in single-digit arithmetic: evidence from velo-cardio-facial syndrome. *Cortex* 45, 177–188. doi: 10.1016/j.cortex.2007.06.003
- De Smedt, B., Swillen, A., Devriendt, K., Frys, J. P., Verschaffel, L., Boets, B., et al. (2008). Cognitive correlates of mathematical disabilities in children with velo-cardio-facial syndrome. *Genet. Couns.* 19, 71–94.
- Debbané, M., Glaser, B., and Eliez, S. (2008). Encoding and retrieval processes in velo-cardio-facial syndrome (VCFS). *Neuropsychology* 22:226. doi: 10.1037/0894-4105.22.2.226
- Dehaene, S. (1997). *The Number Sense: How the Mind Creates Mathematics*. New York, NY: Oxford University Press.
- Dehaene, S. (2007). "Symbols and quantities in parietal cortex: elements of a mathematical theory of number representation and manipulation," in *Sensorimotor Foundations of Higher Cognition - Attention and Performance XXII*, eds P. Haggard, Y. Rossetti, and M. Kawato (Cambridge: Harvard University Press), 527–574.
- Dehaene, S., Izard, I., and Piazza, M. (2005). *Control Over Non-Numerical Parameters in Numerosity Experiments*. Available online at: www.unicog.org/docs/DocumentationDotsGeneration.doc
- Denckla, M. B. (1997). "The neurobehavioral examination in children," in *Behavioral Neurology and Neuropsychology*, eds T. E. Feinberg and M. J. Farah (New York, NY: McGraw-Hill), 721–728.
- Denckla, M. B. (2003). ADHD: topic update. *Brain Dev.* 25, 383–389. doi: 10.1016/S0387-7604(03)00057-3
- Elliott, J. G., and Grigorenko, E. L. (2014). *The Dyslexia Debate (No. 14)*. London: Cambridge University Press.
- Espe, S. (2018). MalaCards: the human disease database. *J. Med. Library Assoc.* 106, 140. doi: 10.5195/JMLA.2018.253
- Fagerberg, C. R., Graakjaer, J., Heinel, U. D., Ousager, L. B., Dreyer, I., Kirchhoff, M., et al. (2012). Heart defects and other features of the 22q11 distal deletion syndrome. *Eur. J. Med. Genet.* 56, 98–107. doi: 10.1016/j.ejmg.2012.09.009
- Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53–72. doi: 10.1016/j.jecp.2014.01.013
- Ferreira, F. O., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krininger, H., Willmes, K., et al. (2012). Explaining school mathematics performance from symbolic and nonsymbolic magnitude processing: similarities and differences between typical and low-achieving children. *Psychol. Neurosci.* 5, 37–46. doi: 10.3922/j.psns.2012.1.06
- Figueiredo, V. L. M. (2002). *WISC-III: Escala de Inteligência Wechsler para Crianças. Manual Adaptação e Padronização Brasileira*. São Paulo: Casa do Psicólogo.
- Figueiredo, V. L. M., and Nascimento, E. (2007). Desempenhos nas duas tarefas do subteste dígitos do WISC-III e do WAIS-III. *Psicologia* 23, 313–318. doi: 10.1590/S0102-37722007000300010
- Geary, D. C., Hoard, M. K., Byrd-Craven, J., and DeSoto, M. C. (2004). Strategy choices in simple and complex addition: Contributions of working memory and counting knowledge for children with mathematical disability. *J. Exp. Child Psychol.* 88, 121–151. doi: 10.1016/j.jecp.2004.03.002
- Genome Reference Consortium (2018). Available online at: <https://www.ncbi.nlm.nih.gov/grc/human/data>
- Gerdes, M., Solot, C., Wang, P. P., McDonald-McGinn, D. M., and Zackai, E. H. (2001). Taking advantage of early diagnosis: preschool children with the 22q11.2 deletion. *Genet. Med.* 3:40. doi: 10.1097/00125817-200101000-00009
- Gomides, M. R. A., Martins, G. Z., Starling-Alves, I., Júlio-Costa, A., Jaeger, A., and Haase, V. G. (2018). Heterogeneity of math difficulties and its implications for interventions in multiplication skills. *Dement. Neuropsychol.* 12, 256–263. doi: 10.1590/1980-57642018dn12-030006
- Gothelf, D., Eliez, S., Thompson, T., Hinard, C., Penniman, L., Feinstein, C., et al. (2005). COMT genotype predicts longitudinal cognitive decline and psychosis in 22q11.2 deletion syndrome. *Nat. Neurosci.* 8:1500. doi: 10.1038/nn1572
- Gothelf, D., Frisch, A., Michaelovsky, E., Weizman, A., and Shprintzen, R. J. (2009). Velocardiofacial syndrome. *J. Ment. Health Res. Intellect. Disabil.* 2, 149–167. doi: 10.1080/19315860902756136
- Green, T., Gothelf, D., Glaser, B., Debbané, M., Frisch, A., Kotler, M., et al. (2009). Psychiatric disorders and intellectual functioning throughout development in velocardiofacial (22q11.2 deletion) syndrome. *J. Am. Acad. Child Adolesc. Psychiatry* 48, 1060–1068. doi: 10.1097/CHI.0b013e3181b76683
- Haase, V. G., Júlio-Costa, A., Lopes-Silva, J. B., Starling-Alves, I., Antunes, A., Pinheiro-Chagas, P., et al. (2014). Contributions from specific and general factors to unique deficits: two cases of mathematics learning difficulties. *Front. Psychol.* 5:102. doi: 10.3389/fpsyg.2014.00102
- Halberda, J., Mazocco, M. M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* 455, 665–668. doi: 10.1038/nature07246
- Hale, J., Alfonso, V., Berninger, V., Bracken, B., Christo, C., Clark, E., et al. (2010). Critical issues in response-to-intervention, comprehensive evaluation, and specific learning disabilities identification and intervention: an expert white paper consensus. *Learn. Disabil. Q.* 33, 223–236. doi: 10.1177/073194871003300310

- Hannonen, R., Komulainen, J., Eklund, K., Tolvanen, A., Riikonen, R., and Ahonen, T. (2010). Verbal and academic skills in children with early-onset type 1 diabetes. *Dev. Med. Child Neurol.* 52, e143–e147. doi: 10.1111/j.1469-8749.2010.03648.x
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *J. Res. Math. Educ.* 21, 33–46.
- Huang, C. J., Chiu, H. J., Lan, T. H., Wang, H. F., Kuo, S. W., Chen, S. F., et al. (2010). Significance of morphological features in schizophrenia of a Chinese population. *J. Psychiatr. Res.* 44, 63–68. doi: 10.1016/j.jpsychires.2009.06.004
- Jacobson, C., Shearer, J., Habel, A., Kane, F., Tsakanikos, E., and Kravariti, E. (2010). Core neuropsychological characteristics of children and adolescents with 22q11.2 deletion. *J. Intell. Disabil. Res.* 54, 701–713. doi: 10.1111/j.1365-2788.2010.01298.x
- Jacobson, J. L., Dodge, N. C., Burden, M. J., Klorman, R., and Jacobson, S. W. (2011). Number processing in adolescents with prenatal alcohol exposure and ADHD: differences in the neurobehavioral phenotype. *Alcoholism* 35, 431–442. doi: 10.1111/j.1530-0277.2010.01360.x
- Johnson, M. H. (2012). Executive function and developmental disorders: the flip side of the coins. *Trends Cogn. Sci.* 16, 454–457. doi: 10.1016/j.tics.2012.07.001
- Júlio-Costa, A., Starling-Alves, I., Lopes-Silva, J. B., Wood, G., and Haase, V. G. (2015). Stable measures of number sense accuracy in math learning disability: Is it time to proceed from basic science to clinical application? *Psych J.* 4, 218–225. doi: 10.1002/pchj.114
- Karagiannakis, G., Baccaglini-Frank, A., and Papadatos, Y. (2014). Mathematical learning difficulties subtypes classification. *Front. Hum. Neurosci.* 8:57. doi: 10.3389/fnhum.2014.00057
- Karayorgou, M., Simon, T. J., and Gogos, J. A. (2010). 22q11.2 microdeletions: linking DNA structural variation to brain dysfunction and schizophrenia. *Nat. Rev. Neurosci.* 11:402. doi: 10.1038/nrn2841
- Klein, E., Nuerk, H. C., Wood, G., Knops, A., and Willmes, K. (2009). The exact vs. approximate distinction in numerical cognition may not be exact, but only approximate: How different processes work together in multi-digit addition. *Brain Cogn.* 69, 369–381. doi: 10.1016/j.bandc.2008.08.031
- Krajcsi, A., Lukács, A., Ignács, J., Racsmány, M., and Pléh, C. (2009). Numerical abilities in Williams syndrome: dissociating the analogue magnitude system and verbal retrieval. *J. Clin. Exp. Neuropsychol.* 31, 439–446. doi: 10.1080/13803390802244126
- Lacerda, S. S. (2012). *Características Psicométricas do Teste de Aprendizagem Auditivo-Verbal de Rey e do Teste de Aprendizagem Visual de Desenhos de Rey para a população brasileira*, Unpublished doctoral thesis. Universidade de São Paulo, São Paulo.
- Lajiness-O'Neill, R., Beaulieu, I., Asamoah, A., Titus, J. B., Bawle, E., Ahmad, S., et al. (2006). The neuropsychological phenotype of velocardiofacial syndrome (VCFs): Relationship to psychopathology. *Arch. Clin. Neuropsychol.* 21, 175–184.
- Landerl, K., Bevan, A., and Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: a study of 8–9-year-old students. *Cognition* 93, 99–125. doi: 10.1016/j.cognition.2003.11.004
- Lee, J. (2009). Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries. *Learn. Individ. Differ.* 19, 355–365. doi: 10.1016/j.lindif.2008.10.009
- Leibovich, T., and Henik, A. (2014). Comparing performance in discrete and continuous comparison tasks. *Q. J. Exp. Psychol.* 67, 899–917. doi: 10.1080/17470218.2013.837940
- Leibovich, T., Katzin, N., Harel, M., and Henik, A. (2017). From “sense of number” to “sense of magnitude”: The role of continuous magnitudes in numerical cognition. *Behav. Brain Sci.* 40:e164. doi: 10.1017/S0140525X16000960
- Levin, H. S., Scheller, J., Rickard, T., Grafman, J., Martinkowski, K., Winslow, M., et al. (1996). Dyscalculia and dyslexia after right hemisphere injury in infancy. *Arch. Neurol.* 53, 88–96. doi: 10.1001/archneur.1996.00550010108024
- Libertus, M. E., Feigenson, L., Halberda, J., and Landau, B. (2014). Understanding the mapping between numerical approximation and number words: Evidence from Williams syndrome and typical development. *Dev. Sci.* 17, 905–919. doi: 10.1111/desc.12154
- Lindgren, V., McRae, A., Dineen, R., Saulsberry, A., Hoganson, G., and Schrift, M. (2015). Behavioral abnormalities are common and severe in patients with distal 22q11.2 microdeletions and microduplications. *Mol. Genet. Genom. Med.* 3, 346–353. doi: 10.1002/mgg3.146
- Lonnemann, J., Linkersdörfer, J., Heselhaus, V., Hasselhorn, M., and Lindberg, S. (2011). Relations between balancing and arithmetic skills in children—Evidence of cerebellar involvement? *J. Neurolinguistics* 24, 592–601. doi: 10.1016/j.jneuroling.2011.02.005
- Lopes-Silva, J. B., Moura, R., Júlio-Costa, A., Geraldi Haase, V., and Wood, G. (2014). Phonemic awareness as a pathway to number transcoding. *Front. Psychol.* 5:13. doi: 10.3389/fpsyg.2014.00013
- Lopes-Silva, J. B., Moura, R., Júlio-Costa, A., Wood, G., Salles, J. F., and Haase, V. G. (2016). What is specific and what is shared between numbers and words? *Front. Psychol.* 7:22. doi: 10.3389/fpsyg.2016.00022
- Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., and Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1–6. *Dev. Sci.* 17, 714–726. doi: 10.1111/desc.12152
- Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *J. Res. Math. Educ.* 30, 520–540.
- Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., and Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: the case of the numerical distance effect. *Acta Psychol.* 134, 154–161. doi: 10.1016/j.actpsy.2010.01.006
- Mazzocco, M. M. (2001). Math learning disability and math LD subtypes: evidence from studies of Turner syndrome, fragile X syndrome, and neurofibromatosis type 1. *J. Learn. Disabil.* 34, 520–533. doi: 10.1177/002221940103400605
- Mazzocco, M. M., Feigenson, L., and Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Dev.* 82, 1224–1237. doi: 10.1111/j.1467-8624.2011.01608.x
- Menon, V., Rivera, S. M., White, C. D., Glover, G. H., and Reiss, A. L. (2000). Dissociating prefrontal and parietal cortex activation during arithmetic processing. *Neuroimage* 12, 357–365. doi: 10.1006/nimg.2000.0613
- Merkley, R., Thompson, J., and Scerif, G. (2016). Of huge mice and tiny elephants: exploring the relationship between inhibitory processes and preschool math skills. *Front. Psychol.* 6:1903. doi: 10.3389/fpsyg.2015.01903
- Mikhail, F. M., Burnside, R. D., Rush, B., Ibrahim, J., Godshalk, R., Rutledge, S. L., et al. (2014). The recurrent distal 22q11.2 microdeletions are often de novo and do not represent a single clinical entity: a proposed categorization system. *Genet. Med.* 16:92. doi: 10.1038/gim.2013.79
- Mikhail, F. M., Descartes, M., Piotrowski, A., Andersson, R., Diaz de Ståhl, T., Komorowski, J., et al. (2007). A previously unrecognized microdeletion syndrome on chromosome 22 band q11.2 encompassing the BCR gene. *Am. J. Med. Genet.* 143A, 2178–2184. doi: 10.1002/ajmg.a.31882
- Moberg, P. J., Richman, M. J., Roalf, D. R., Morse, C. L., Graefe, A. C., Brennan, L., et al. (2018). Neurocognitive functioning in patients with 22q11.2 deletion syndrome: a meta-analytic review. *Behav. Genet.* 48, 1–12. doi: 10.1007/s10519-018-9903-5
- Molck, M. C., Vieira, T. P., Sgardioli, I. C., Simioni, M., Dos Santos, A. P., Souza, J., et al. (2013). Atypical copy number abnormalities in 22q11.2 region: Report of three cases. *Eur. J. Med. Genet.* 56, 515–520. doi: 10.1016/j.ejmg.2013.07.002
- Moura, R., Lopes-Silva, J. B., Vieira, L. R., Paiva, G. M., de Almeida Prado, A. C., Wood, G., et al. (2015). From “Five” to 5 for 5 minutes: arabic number transcoding as a short, specific, and sensitive screening tool for mathematics learning difficulties. *Arch. Clin. Neuropsychol.* 30, 88–98. doi: 10.1093/arclin/acu071
- Moura, R. J., Wood, G., Pinheiro-Chagas, P., Lonnemann, J., Krinzinger, H., Willmes, K., et al. (2013). Transcoding abilities in typical and atypical mathematics achievers: the role of working memory, procedural and lexical competencies. *J. Exp. Child Psychol.* 116, 707–727. doi: 10.1016/j.jecp.2013.07.008
- Newbern, J., Zhong, J., Wickramasinghe, R. S., Li, X., Wu, Y., Samuels, I., et al. (2008). Mouse and human phenotypes indicate a critical conserved role for ERK2 signaling in neural crest development. *Proc. Natl. Acad. Sci. U.S.A.* 105, 17115–17120. doi: 10.1073/pnas.0805239105
- Nicolson, R. I., and Fawcett, A. (2010). *Dyslexia, Learning, and the Brain*. Cambridge: MIT Press.
- Nieder, A., and Dehaene, S. (2009). Representation of number in the brain. *Annu. Rev. Neurosci.* 32, 185–208. doi: 10.1146/annurev.neuro.051508.135550
- Nobile, M., Rusconi, M., Bellina, M., Marino, C., Giorda, R., Carlet, O., et al. (2010). COMT Val158Met polymorphism and socioeconomic status interact to predict attention deficit/hyperactivity problems in children aged 10–14. *Eur. Child Adolesc. Psychiatry* 19, 549–557. doi: 10.1007/s00787-009-0080-1

- Oliveira, L. F., Santos, A. O., Vianna, G. S., Di Ninno, C. Q., Giacheti, C. M., Carvalho, M. R., et al. (2014). Impaired acuity of the approximate number system in 22q11.2 microdeletion syndrome. *Psychol. Neurosci.* 7:151. doi: 10.3922/j.psns.2014.02.04
- Oliveira, M., Rigoni, M., Andretta, I., and Moraes, J. F. (2004). Validação do teste figuras complexas de Rey para a população brasileira. *Avaliação Psicológica* 3, 33–38.
- Oliveira-Ferreira, F., Costa, D. S., Micheli, L. R., Oliveira, L. F. S., Pinheiro-Chagas, P., and Haase, V. G. (2012). School Achievement Test: normative data for a representative sample of elementary school children. *Psychol. Neurosci.* 5, 157–164. doi: 10.3922/j.psns.2012.2.05
- Pennington, B. F. (2006). From single to multiple deficit models of developmental disorders. *Cognition* 101, 385–413. doi: 10.1016/j.cognition.2006.04.008
- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012
- Piazza, M., Izard, V., Pinel, P., LeBihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human parietal cortex. *Neuron* 44, 547–555. doi: 10.1016/j.neuron.2004.10.014
- Pinheiro-Chagas, P., Wood, G., Knops, A., Krininger, H., Lonnemann, J., Starling-Alves, I., et al. (2014). In how many ways is the approximate number system associated with exact calculation? *PLoS ONE* 9:e111155. doi: 10.1371/journal.pone.0111155
- Poole, J. L., Burtner, P. A., Torres, T. A., McMullen, C. K., Markham, A., Marcum, M. L., et al. (2005). Measuring dexterity in children using the Nine-hole Peg Test. *J. Hand Ther.* 18, 348–351. doi: 10.1197/j.jht.2005.04.003
- Prado, J. (2018). “The interplay between learning arithmetic and learning to read: insights from developmental cognitive neuroscience,” in *Heterogeneity of Function in Numerical Cognition*, eds A. Henik and W. Fias (Cambridge: Academic Press), 27–49.
- Price, G. R., Palmer, D., Battista, C., and Ansari, D. (2012). Nonsymbolic numerical magnitude comparison: Reliability and validity of different task variants and outcome measures, and their relationship to arithmetic achievement in adults. *Acta Psychol.* 140, 50–57. doi: 10.1016/j.actpsy.2012.02.008
- Raghubar, K. P., Barnes, M. A., and Hecht, S. A. (2010). Working memory and mathematics: a review of developmental, individual difference, and cognitive approaches. *Learn. Individ. Differ.* 20, 110–122. doi: 10.1016/j.lindif.2009.10.005
- Robin, N. H., and Shprintzen, R. J. (2005). Defining the clinical spectrum of deletion 22q11.2. *J. Pediatr.* 147, 90–96. doi: 10.1016/j.jpeds.2005.03.007
- Rocha, M. M., Rescorla, L. A., Emerich, D. R., Silveira, E. F. M., Borsa, J. C., Araújo, L. G. S., et al. (2012). Behavioural/emotional problems in Brazilian children: findings from parents' reports on the child behavior checklist. *Epidemiol. Psychiatr. Sci.* 22, 329–338. doi: 10.1017/S2045796012000637
- Rodningen, O. K., Prescott, T., Eriksson, A. S., and Rosby, O. (2008). 1.4 Mb recurrent 22q11.2 distal deletion syndrome, two new cases expand the phenotype. *Eur. J. Med. Genet.* 51, 646–650. doi: 10.1016/j.ejmg.2008.07.007
- Rousselle, L., and Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: a comparison of symbolic vs non-symbolic number magnitude processing. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005
- Rubinsten, O., and Henik, A. (2009). Developmental dyscalculia: heterogeneity might not mean different mechanisms. *Trends Cogn. Sci.* 13, 92–99. doi: 10.1016/j.tics.2008.11.002
- Rutter, M., Moffitt, T. E., and Caspi, A. (2006). Gene-environment interplay and psychopathology: multiple varieties but real effects. *J. Child Psychol. Psychiatry* 47, 226–261. doi: 10.1111/j.1469-7610.2005.01557.x
- Saitta, S. C., McGrath, J. M., Mensch, H., Shaikh, T. H., Zackai, E. H., and Emanuel, B. S. (1999). A 22q11.2 deletion that excludes UFD1L and CDC45L in a patient with conotruncal and craniofacial defects. *Am. J. Hum. Genet.* 65, 562–566. doi: 10.1086/302514
- Santos, F. H., and Bueno, O. F. A. (2003). Validation of the Brazilian children's test of pseudoword repetition in Portuguese speakers aged 4 to 10 years. *Braz. J. Med. Biol. Res.* 36, 1633–1547. doi: 10.1590/S0100-879X2003001100012
- Santos, F. H., Mello, C. B., Bueno, O. F. A., and Dellatolas, G. (2005). Cross-cultural differences for three visual memory tasks in Brazilian children. *Percept. Mot. Skills* 101, 421–433. doi: 10.2466/pms.101.2.421-433
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372
- Schoch, K., Harrell, W., Hooper, S. R., Ip, E. H., Saldana, S., Kwapil, T. R., et al. (2014). Applicability of the nonverbal learning disability paradigm for children with 22q11.2 deletion syndrome. *J. Learn. Disabil.* 47, 153–166. doi: 10.1177/0022219412443556
- Shalev, R. S., and Gross-Tsur, V. (1993). Developmental dyscalculia and medical assessment. *J. Learn. Disabil.* 26, 134–137. doi: 10.1177/002221949302600206
- Siegler, R. S., and Braithwaite, D. W. (2017). Numerical development. *Annu. Rev. Psychol.* 68, 187–213. doi: 10.1146/annurev-psych-010416-044101
- Simon, T. J. (2008). A new account of the neurocognitive foundations of impairments in space, time, and number processing in children with chromosome 22q11.2 deletion syndrome. *Dev. Disabil. Res. Rev.* 14, 52–58. doi: 10.1002/ddrr.8
- Simon, T. J., Bearden, C. E., Mc-Ginn, D. M., and Zackai, E. (2005a). Visuospatial and numerical cognitive deficits in children with chromosome 22q11.2 deletion syndrome. *Cortex* 41, 145–155. doi: 10.1016/S0010-9452(08)70889-X
- Simon, T. J., Ding, L., Bish, J. P., McDonald-McGinn, D. M., Zackai, E. H., and Gee, J. (2005b). Volumetric, connective, and morphologic changes in the brains of children with chromosome 22q11.2 deletion syndrome: an integrative study. *Neuroimage* 25, 169–180. doi: 10.1016/j.neuroimage.2004.11.018
- Simpson, N. H., Addis, L., Brandler, W. M., Slonims, V., Clark, A., Watson, J., et al. (2014). Increased prevalence of sex chromosome aneuploidies in specific language impairment and dyslexia. *Dev. Med. Child Neurol.* 56, 346–353. doi: 10.1111/dmnc.12294
- Smets, K., Sasanguie, D., Szűcs, D., and Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *J. Cogn. Psychol.* 27, 310–325. doi: 10.1080/20445911.2014.996568
- Spinelli-Silva, S., Bispo, L. M., Gil-da-Silva-Lopes, V. L., and Vieira, T. P. (2017). Distal deletion at 22q11.2 as differential diagnosis in Craniofacial Microsomia: case report and literature review. *Eur. J. Med. Genet.* 61, 262–268. doi: 10.1016/j.ejmg.2017.12.013
- Stankov, L., Lee, J., Luo, W., and Hogan, D. J. (2012). Confidence: A better predictor of academic achievement than self-efficacy, self-concept and anxiety? *Learn. Individ. Differ.* 22, 747–758. doi: 10.1016/j.lindif.2012.05.013
- Stein, L. M. (1994). *Teste de Desempenho Escolar: Manual para Aplicação e Interpretação*. São Paulo: Casa do Psicólogo.
- Swanson, H. L., and Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: both executive and phonological processes are important. *J. Exp. Child Psychol.* 79, 294–321. doi: 10.1006/jecp.2000.2587
- Swillen, A., Devriendt, K., Legius, E., Eyskens, B., Dumoulin, M., Gewillig, M., et al. (1997). Intelligence and psychosocial adjustment in velocardiofacial syndrome: a study of 37 children and adolescents with VCFS. *J. Med. Genet.* 34, 453–458. doi: 10.1136/jmg.34.6.453
- Swillen, A., Van deputte, L., Cracco, J., Maes, B., Ghesquière, P., Devriendt, K., et al. (1999). Neuropsychological, learning and psychosocial profile of primary school aged children with the velo-cardio-facial syndrome (22q11 deletion): evidence for a nonverbal learning disability? *Child Neuropsychol.* 5, 230–241. doi: 10.1076/0929-7049(199912)05:04;1-R;FT230
- Szűcs, D., Devine, A., Soltesz, F., Nobes, A., and Gabriel, F. (2013). Developmental dyscalculia is related to visuo-spatial memory and inhibition impairment. *Cortex* 49, 2674–2688. doi: 10.1016/j.cortex.2013.06.007
- Tan, T. Y., Collins, A., James, P. A., McGillivray, G., Stark, Z., Gordon, C. T., et al. (2011). Phenotypic variability of distal 22q11.2 copy number abnormalities. *Am. J. Med. Genet. Part A* 155, 1623–1633. doi: 10.1002/ajmg.a.34051
- Tenenbaum, A., Talia Dor, M. D., Yael Castiel, R. N., and Sapir, A. (2011). Fetal alcohol spectrum disorder in Israel: increased prevalence in an at-risk population. *Religion* 59:9.
- Van Ijzendoorn, M. H., Juffer, F., and Poelhuis, C. W. K. (2005). Adoption and cognitive development: a meta-analytic comparison of adopted and nonadopted children's IQ and school performance. *Psychol. Bull.* 131:301. doi: 10.1037/0033-2909.131.2.301
- Vandervort, L. (2017). The origin of mathematics and number sense in the cerebellum: with implications for finger counting and dyscalculia. *Cerebell. Ataxias* 4:12. doi: 10.1186/s40673-017-0070-x

- Vaz, I. A., Cordeiro, P. M., de Macedo, E. C., and Lukasova, K. (2010). Memória de trabalho em crianças avaliada pela tarefa de Brown-Peterson. *Pró-Fono* 22, 95–100. doi: 10.1590/S0104-56872010000200005
- Verhoeven, W., Egger, J., Brunner, H., and De Leeuw, N. (2011). A patient with a de novo distal 22q11.2 microdeletion and anxiety disorder. *Am. J. Med. Genet. A* 155, 392–397. doi: 10.1002/ajmg.a.33802
- Vicari, S., Mantovan, M., Addona, F., Costanzo, F., Verucci, L., and Menghini, D. (2011). Neuropsychological profile of Italian children and adolescents with 22q11.2 deletion syndrome with and without intellectual disability. *Behav. Genet.* 42, 287–298. doi: 10.1007/s10519-011-9499-5
- Villalon-Reina, J., Jahanshad, N., Beaton, E., Toga, A. W., Thompson, P. M., and Simon, T. J. (2013). White matter microstructural abnormalities in girls with chromosome 22q11. 2 deletion syndrome, Fragile X or Turner syndrome as evidenced by diffusion tensor imaging. *Neuroimage* 81, 441–454. doi: 10.1016/j.neuroimage.2013.04.028
- Wilson, A. J., and Dehaene, S. (2007). Number sense and developmental dyscalculia. *Hum. Behav. Learn. Dev. Brain Atypical Dev.* 2, 212–237.
- Wong, T. T. Y., Ho, C. S. H., and Tang, J. (2014). Identification of children with mathematics learning disabilities (MLDs) using latent class growth analysis. *Res. Dev. Disabil.* 35, 2906–2920. doi: 10.1016/j.ridd.2014.07.015
- Woodin, M., Wang, P. P., Aleman, D., Donald-McGinn, D., Zackai, E., and Moss, E. (2001). Neuropsychological profile of children and adolescents with the 22q11.2 microdeletion. *Genet. Med.* 3, 34–39. doi: 10.1097/00125817-200101000-00008
- Xu, J., Fan, Y. S., and Siu, V. M. (2008). A child with features of Goldenhar syndrome and a novel 1.12 Mb deletion in 22q11.2 by cytogenetics and oligonucleotide array CGH: Is this a candidate region for the syndrome? *Am. J. Med. Genet. A* 146A, 1886–1889. doi: 10.1002/ajmg.a.32359
- Zeit, M. J., Lerner, P. P., Ay, F., Van Nostrand, E., Heidmann, J. D., Noble, W. S., et al. (2013). Implications of COMT long-range interactions on the phenotypic variability of 22q11. 2 deletion syndrome. *Nucleus* 4, 487–493. doi: 10.4161/nucl.27364
- Zoukoku, K., and Temple, C. M. (2016). The processing of number scales beyond whole numbers in development: dissociations in arithmetic in Turner's syndrome. *Cogn. Neuropsychol.* 33, 277–298. doi: 10.1080/02643294.2016.1179178

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Oliveira, Júlio-Costa, Santos, Carvalho and Haase. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Approximate Arithmetic Training Improves Informal Math Performance in Low Achieving Preschoolers

Emily Szkudlarek* and Elizabeth M. Brannon

Department of Psychology, University of Pennsylvania, Philadelphia, PA, United States

Recent studies suggest that practice with approximate and non-symbolic arithmetic problems improves the math performance of adults, school aged children, and preschoolers. However, the relative effectiveness of approximate arithmetic training compared to available educational games, and the type of math skills that approximate arithmetic targets are unknown. The present study was designed to (1) compare the effectiveness of approximate arithmetic training to two commercially available numeral and letter identification tablet applications and (2) to examine the specific type of math skills that benefit from approximate arithmetic training. Preschool children ($n = 158$) were pseudo-randomly assigned to one of three conditions: approximate arithmetic, letter identification, or numeral identification. All children were trained for 10 short sessions and given pre and post tests of informal and formal math, executive function, short term memory, vocabulary, alphabet knowledge, and number word knowledge. We found a significant interaction between initial math performance and training condition, such that children with low pretest math performance benefited from approximate arithmetic training, and children with high pretest math performance benefited from symbol identification training. This effect was restricted to informal, and not formal, math problems. There were also effects of gender, socio-economic status, and age on post-test informal math score after intervention. A median split on pretest math ability indicated that children in the low half of math scores in the approximate arithmetic training condition performed significantly better than children in the letter identification training condition on post-test informal math problems when controlling for pretest, age, gender, and socio-economic status. Our results support the conclusion that approximate arithmetic training may be especially effective for children with low math skills, and that approximate arithmetic training improves early informal, but not formal, math skills.

Keywords: preschool math, approximate number system, cognitive training, approximate arithmetic, numerical cognition, tablet application

OPEN ACCESS

Edited by:

Ann Dowker,
University of Oxford, United Kingdom

Reviewed by:

Paige H. Fisher,
Seton Hall University, United States
Annemie Desoete,
Ghent University, Belgium

*Correspondence:

Emily Szkudlarek
emilysz@sas.upenn.edu

Specialty section:

This article was submitted to
Developmental Psychology,
a section of the journal
Frontiers in Psychology

Received: 09 November 2017

Accepted: 10 April 2018

Published: 15 May 2018

Citation:

Szkudlarek E and Brannon EM (2018)
Approximate Arithmetic Training
Improves Informal Math Performance
in Low Achieving Preschoolers.
Front. Psychol. 9:606.
doi: 10.3389/fpsyg.2018.00606

INTRODUCTION

Early math competency is an important predictor of later academic achievement and a variety of measures of adult health and economic well-being (Duncan et al., 2007; Jordan et al., 2009, 2010; Reyna et al., 2009; Geary et al., 2013; Gerardi et al., 2013). It is critical that children enter kindergarten and first grade prepared to embark on formal math learning, however, there is wide variation in the level of math skill children acquire during the preschool years (Jordan et al., 2006). Conceptual knowledge of addition and subtraction is an

especially important skill for children at the beginning of formal math education (Nunes et al., 2007; Ching and Nunes, 2017). Therefore, improving early conceptual knowledge of arithmetic is an important way to enhance math readiness in preschool children.

The Approximate Number System (ANS) supports an intuitive sense of number that allows adults, human infants, and many non-human animals to compare, estimate, and manipulate non-symbolic and approximate numerical quantities (Feigenson et al., 2004). For example, the ANS allows children to distinguish which of two sets of objects is greater in number. There is a modest but significant relation between ANS acuity and symbolic math skills (see Chen and Li, 2014; Fazio et al., 2014; Schneider et al., 2016 for meta-analyses). Specifically, children and adults with greater ANS acuity score better on math achievement measures such as the TEMA, the calculation portion of the Woodcock Johnson, or even self-reported SAT exams (Halberda et al., 2008, 2012). This relation suggests that the ANS may be a building block upon which children anchor their concept of symbolic number. Previous research has demonstrated that children can solve math problems non-symbolically and approximately before they comprehend the same operations symbolically (Barth et al., 2005). With the ANS, young children can compare, add, subtract, multiply, and divide, and solve simple linear equations using sets of objects with ratio-dependent precision (Barth et al., 2006; McCrink and Spelke, 2010, 2016; Kibbe and Feigenson, 2015). In contrast to these prodigious non-symbolic and approximate mathematical abilities, children must be explicitly taught how to solve the same symbolic mathematical problems effectively over years of formal schooling.

To further test the hypothesis that ANS representations serve as a building block for symbolic mathematics, recent work has tested the possible causal relation between ANS based tasks and symbolic math skills. In the first of these studies, Park and Brannon (2013, 2014) trained adults on an approximate arithmetic task and tested their symbolic arithmetic fluency before and after training. During approximate arithmetic training, subjects watched addition and subtraction events depicted with animated arrays of dots. For example, during an addition trial, an array of dots appeared and then moved behind an opaque box. A second array of dots then appeared and also moved behind the box. After watching this animation, the subject imagines the sum behind the box and compares this imagined quantity to a second visible quantity. Adults trained on this approximate arithmetic task showed greater improvement on a symbolic arithmetic assessment compared to a no contact control group, a group trained on general knowledge facts, a group trained to rapidly order numerals, a group trained on a visuo-spatial short term memory task, and a group trained on approximate numerosity comparisons. Thus, for adults, practice mentally manipulating approximate quantities in arithmetic operations yielded a benefit for symbolic arithmetic performance that was not afforded by any of the control training tasks. This finding raised the important question as to whether non-symbolic and approximate arithmetic training could also be effective for children. If shown to be effective for

preschoolers, approximate arithmetic training could be a useful tool for introducing arithmetic concepts to children before they are ready to master symbolic arithmetic in the classroom.

A handful of experiments have explored this possibility by training children on approximate arithmetic tasks and testing their symbolic math abilities after training (Hyde et al., 2014; Khanum et al., 2016; Park et al., 2016; Dillon et al., 2017). Hyde et al. (2014) found that first grade children who completed a session of approximate arithmetic or dot comparison training were faster at completing a symbolic arithmetic test than children who had completed a training session of line length addition or brightness comparison. This finding was replicated in an independent sample of children, suggesting that approximate arithmetic training improves arithmetic fluency (Khanum et al., 2016). In a large scale study conducted in India, approximate arithmetic combined with geometry training improved non-symbolic but not symbolic math performance in preschool and elementary school children (Dillon et al., 2017). Children who participated in the non-symbolic math training condition maintained higher non-symbolic math skills 1 year after training compared to the children in the control group. Park et al. (2016) tested the effectiveness of approximate arithmetic training with preschool children using a pre/post test training paradigm. An approximate arithmetic tablet application called Max's Math Game was created to mirror the adult approximate arithmetic training studies of Park and Brannon (2013, 2014). Over 10 training sessions preschool aged children played Max's Math Game or a non-math picture-memory game. Children were tested with The Third Edition of the Test of Early Mathematics Achievement (TEMA-3; Ginsburg and Baroody, 2003), and with measures of vocabulary, short term memory, and executive function before and after training. Preschoolers who trained on the approximate arithmetic task selectively improved on the TEMA-3 significantly more than children who trained on the picture-memory game. Taken together, the research on non-symbolic math training suggests that practice with approximate and non-symbolic arithmetic may be an effective way to improve the math skills of young children (but see Szucs and Myers, 2017).

The current study aims to advance approximate arithmetic training research in two ways. First, the current study was designed to provide insight into the nature of the symbolic math skills that approximate arithmetic training benefits. Prior research has found that ANS acuity correlates with TEMA-3 questions that assess informal, but not formal, math abilities (Libertus et al., 2013). Thus, it is possible that approximate arithmetic training selectively improves informal, but not formal, symbolic math abilities. Informal math abilities include counting, assessments of numerical magnitude, and knowledge of the ordinal relationship between numbers in the counting sequence, while formal math abilities include fact retrieval and numeral identification (Ginsburg and Baroody, 2003; Jordan et al., 2009). Informal symbolic math skills require children to use number words and symbols in mathematical operations. For example, the informal math question "You have 4 pennies. I give you 2 more pennies. How many pennies do you have altogether?" is a conceptual test of addition. In contrast, formal math skills involve the memorization of math facts. For example, when a child is

shown the numeral “4” and asked “What number is this?” the child must recall that the symbol “4” corresponds to the word “four.” During approximate arithmetic training children do not gain experience with the formal math skill of identifying that the symbol “4” corresponds to the word “four,” however, the process of addition is modeled repeatedly. Thus, approximate and non-symbolic practice with addition and subtraction may induce improved performance selectively on informal math problems that test knowledge of arithmetic concepts. To test this hypothesis in the current study, we created a measure of early math skills inspired by the Number Sense Screener (NSS; Research Edition: Glutting and Jordan, 2012). Many standardized tests of math for young children, like the TEMA-3, are good measures of general early math performance, but due to age standardization and titration procedures it is difficult to break down the specific math skills improved by training. Our measure is split into sections, with each section defined by a specific math skill. This design allowed us to separately evaluate improvements in informal and formal math skills as a result of approximate arithmetic training.

The second aim of the current study was to compare the effectiveness of approximate arithmetic training to existing math educational practices. Specifically, we compared approximate arithmetic training to two commercially available applications designed to improve symbol knowledge, the 123 Ninja and ABC Ninja games (alligatorapps.com). Previous studies have compared the effectiveness of approximate arithmetic training to control groups trained with non-numerical tasks, and not to educationally relevant math games. For approximate arithmetic to be useful in a classroom, it should be at least as effective as other age appropriate math games. In the control training games used in the current study, children see multiple numerals (123 Ninja) or letters (ABC Ninja) floating across the screen. The child then hears one letter or one number word and is tasked with selecting the appropriate symbol. Educational tablet applications have gained popularity in recent years, but they have been largely untested for their actual educational outcomes (Hirsh-Pasek et al., 2015). We included the 123 Ninja game to assess whether age appropriate symbolic math training would be as effective at improving math performance as approximate arithmetic training. We also included the ABC Ninja game to provide an active control condition that measures the baseline effects of playing any educational tablet application with an experimenter.

Overall, our design allows for the comparison of approximate arithmetic training to educationally relevant control conditions, and can determine with greater specificity the type of math skills improved due to approximate arithmetic training. Our approximate arithmetic training application, Max’s Math Game, has been shown to improve early math skills as measured by the TEMA-3, but ANS acuity correlates with the informal but not formal math questions on the TEMA-3 (Libertus et al., 2013; Park et al., 2016). Moreover, approximate arithmetic training does not involve practice with formal math skills. These facts led us to predict that approximate arithmetic training would improve informal, but not formal, math skills. Conversely, we predicted that 123 Ninja, a numeral identification training application, would improve the formal skill of numeral recognition. Finally,

we predicted that letter identification training (ABC Ninja) would not improve either formal or informal math skills, but would improve alphabet knowledge. Finally, consistent with the findings of Park et al. (2016), we predicted no effect of training condition on vocabulary, executive function, or short term memory.

METHODS

Participants

One hundred and fifty-eight children with a median age of 4.68 (3.27–5.72) were pseudo-randomly assigned to one of three conditions to minimize differences at pretest in age, sex, PPVT, and math score across the groups. Written parental consent was collected in accordance with a protocol accepted by Duke University’s Institutional Review Board. Children were drawn from 7 different preschools and we attempted to consent all parents with children aged 3–5 at each preschool location. Five of the 7 preschools were in the North Carolina Pre-K program. This program provides preschool education for children of low socioeconomic status. In order to be eligible for this program, parental income must be no more than 75% of the state median income. Eighty-four percent of the participants in our study were enrolled in the NC-PreK program. We obtained detailed demographic data for 86 children. Among this subset of our sample, 26% identified as Hispanic, 63% as not Hispanic, and 11% did not report. Sixty-one percent of the sample identified as African American, 6% as Caucasian, 9% as Asian, American Indian or mixed race, and 24% did not report. Thirty-four percent of the mothers reported a high school degree or some high school, 38% reported a college degree or some college, 16% graduate degree or some graduate school, 3% technical school degree, and 9% chose not to report. Seventeen additional participants were consented but did not complete the study due to a variety of reasons including leaving the school, attending the school on a limited basis, family vacation, or turning 6 years old before testing began. One participant who completed the study was excluded from analysis due to frequent absences and completing the post-test session after an extended winter break (111 days between pre and post tests).

Procedure

Participants completed a total of 14 experimental sessions: 2 pre-test sessions, 10 training sessions, and 2 post-test sessions. All sessions were administered in a quiet location at the preschool. Each pre and post-test session lasted between 20 and 40 min, and was administered individually. The experimenter who administered pre and post testing was blind to the condition of the child, except for the first 9 participants tested. Pre and post tests consisted of a symbolic math test based on the Number Sense Screener (NSS; Research Edition: Glutting and Jordan, 2012), a short-term memory task, a Stroop interference task, a standard dimensional card sorting task, the Peabody Picture Vocabulary Test 4th Edition (PPVT-4; Dunn and Dunn, 1997), an alphabet knowledge task, and the Give-a-Number task (Wynn, 1990, 1992). Each assessment had two versions and each child was given a different version of the test for pre and post testing

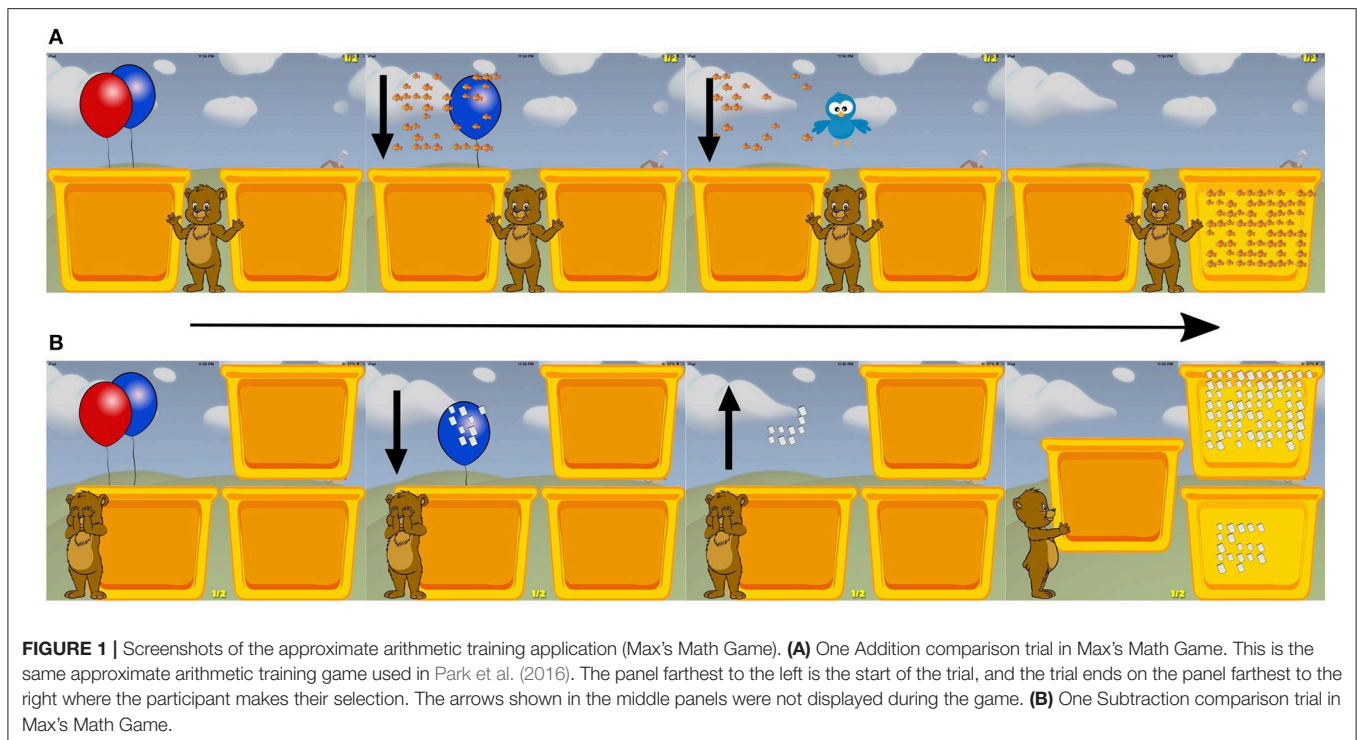


FIGURE 1 | Screenshots of the approximate arithmetic training application (Max's Math Game). **(A)** One Addition comparison trial in Max's Math Game. This is the same approximate arithmetic training game used in Park et al. (2016). The panel farthest to the left is the start of the trial, and the trial ends on the panel farthest to the right where the participant makes their selection. The arrows shown in the middle panels were not displayed during the game. **(B)** One Subtraction comparison trial in Max's Math Game.

with the order of versions counterbalanced across participants. The median time between pre and post test was 27 days. Training sessions occurred in small groups of 3–8 children. During the first training session, children were instructed in how to play the game in detail. After the first training session, children were instructed as needed. Children were monitored for the full 12 min of training to ensure the game was played properly and with full attention. Children wore headphones during training to increase attention to the verbal instructions in each game. Children were rewarded after each experimental session with a sticker of their choice. After all the children in a classroom had completed all 14 sessions, each child received an educational book and building toy, and the classroom was given an additional educational gift chosen by the teacher.

Training Tasks

Approximate Arithmetic Training (Max's Math)

A trial began with Max (a cartoon bear) holding a red balloon (**Figures 1A,B**). Children were instructed to “pop the red balloon,” by touching it, at which point the balloon popped and dropped an array of 4–64 discrete objects (e.g., ears of corn, elephants) into an opaque container. There were four trial types: Addition Comparison, Subtraction Comparison, Matching Addition, and Matching Subtraction. During the addition trials, a second blue balloon popped and dropped more of the objects into the same container. On subtraction trials, the blue balloon popped to reveal a bird that flew in and removed a portion of the original set of objects from the container and off the screen. On comparison trials, children compared the remembered sum or difference to a new target array that appeared in a second container to the right, and were instructed to choose the

container that held more items. On matching trials, children were shown two new target arrays with visible objects, and children picked the container that held the same number of items as the remembered sum or difference. Children were given each of the 4 trial types in separate 10-trial blocks. After two blocks of 10 trials each, a short 45–60 s movie played to maintain attention. On half of the matching trials the container with the smaller number of items was the correct choice. Children completed as many trials as possible in 12 min. The median number of trials completed per session was 39 (standard deviation of 5.3) or about 1 block of each trial type per 12-min training session.

Difficulty was titrated based on performance by manipulating the ratio of the target array to the remembered sum or difference. To do this we varied the numerical distance between the target and the alternative in a log-base 2 scale (the log difference level). The game began with a log difference level of 2 (the ratio between the arrays was $1:2^2$ or $1:4$). For example, if the target was 20, the alternative was either (20×4) or $(20/4)$. The log difference level changed based on the child's average accuracy in a block of 10 trials. If the average accuracy was $<60\%$ the log difference level increased by one of the values randomly chosen from [0.08, 0.09, 0.10, 0.11, 0.12] for the next block. If the average accuracy was between 65 and 80% the log difference level stayed the same. If the average accuracy for the block was greater or equal to 80%, the log difference level decreased by one of the values randomly chosen from [0.13, 0.14, 0.15, 0.16, 0.17] for the next block. Each trial type was titrated separately. The log difference level was never allowed to exceed 2.

123 Ninja—Numeral Identification Training

123 Ninja is a commercial educational application found on the Apple App Store, and is made by Alligator Apps

(alligatorapps.com). In this game children hear a number word, as two or three numerals appear on the screen. Children must swipe the numeral corresponding to the number word they hear with their finger. If they correctly identify the numeral, the game makes a sound indicating a correct response, and the bar at the top of the screen begins to fill up. Once the bar is filled completely, the child is awarded a star, which then appears at the top of the screen throughout the rest of the session. If a child swipes the incorrect numeral, a popping noise is made and the incorrectly swiped numeral turns gray. The same number is repeated until a child swipes the correct numeral. The task was not titrated for difficulty. Children completed as many trials as they could in the 12-min training session. The numerals ranged from 0 to 19. Each number was identified ~ 3 times over the course of 1 training session.

ABC Ninja—Letter Identification Training

ABC Ninja is made by the same app developer, Alligator Apps. It is exactly the same as 123 Ninja, except that letters appear on the screen instead of numerals. All capital letters A-Z were used.

Pre and Post Tests

Informal Math Test

As a measure of symbolic math we modified the NSS to make it appropriate for preschoolers (NSS; Research Edition: Glutting and Jordan, 2012). We used this measure instead of the TEMA-3, because the NSS is divided into question types. This allowed us to assess performance on informal and formal math questions separately. Our test included five informal problem types: Counting, Symbolic Number Comparisons, Nonverbal Calculation, Arithmetic Story Problems, and Simple Arithmetic Problems. The problems used in the NSS were expanded, the wording of some problems changed, and a B version of the test was created to make the test appropriate for preschool aged children and our research questions. The counting section included counting items on a page, and verbally counting as high as possible. The symbolic number comparisons section included questions such as “Which is bigger or more, 6 or 8?” and “What number comes right after 7?” with visual displays of the numerals. In nonverbal calculation, children were shown 1–4 tokens that were subsequently moved under an opaque paper flashcard in an arithmetic operation. For example, on one trial a child was first shown 3 tokens which were then hidden under the flashcard. Then, the child was shown 2 new tokens which were then hidden under the same flashcard. The child had to put the exact same number of tokens under their own flashcard to match the answer to the addition or subtraction problem modeled by the experimenter. In the arithmetic story problems section there were questions such as “You have 4 pennies. I give you 2 more pennies. How many pennies do you have altogether?” The simple arithmetic problem section included questions such as “How much is 7 take away 4?” and “How much is 2 and 1 altogether?” while the numerals in the question appeared on a book in front of the child. There were 28 total questions, and performance was measured as the total number of correctly answered questions. The published test re-retest reliability score for the NSS is 0.81 for kindergarteners measured a month apart. Additionally, we

correlated pre and post test scores of all subjects to get a proxy measure of reliability in our sample. The Pearson correlation coefficient between the pre and post test scores of the informal math test was 0.65, indicating reasonable reliability.

Formal Math Test

The formal section of the math test consisted of 8 numeral identification questions. Children were shown a numeral and asked “What number is this?” Performance was measured as the total number of correct answers. The correlation coefficient between pre and post test formal math score was 0.81, indicating high reliability.

Number Word Knowledge

The cardinality section was the give-a-number task (Wynn, 1990, 1992). In this task, each child was presented with a plate of fish, introduced to a stuffed dinosaur, and told the animal was hungry. The experimenter then asked “Can you give the dinosaur one fish?” Once the child placed fish on the plate she/he was asked “Is that one fish?” Children were allowed to fix their responses, and there was no time limit. If successful, the child was then asked to give the dinosaur two fish and given time to correct their answer. On each subsequent trial children were asked to give the dinosaur $N+1$ (if successful) or $N-1$ (if unsuccessful) fish. No feedback was provided. Trials continued until there were 2 successes at a given N and two failures at $N-1$, with $N = 6$ as the maximum value requested. Children were categorized by knower level defined as the highest number they could successfully produce. The correlation coefficient between pretest knower level and post-test knower level was 0.77.

Short Term Memory Task: Letter Span

Children listened as the experimenter read a string of letters. The child was then asked to repeat the letters back in the same order. There were 6 blocks of 5 trials each. In each successive block the string of letters increased by one letter, so that the first block contained two letter strings and the last block contained seven letters. Children continued until they missed 3 or more trials in one block. For the A and B versions of the task the same letters were used, but in a different order. Only monosyllabic letters were used, and letters with similar sounds (e.g., v and b) were excluded. We used this short term memory task for consistency with the Park et al. (2016) experiment. However, it is important to note this is a measure of verbal short term memory, not visual short term memory. One participant in the ABC Ninja condition did not complete this task. Performance was measured as the total number of successful trials. The correlation coefficient between pretest and post-test short term memory score was 0.75.

Executive Function: Standard Dimensional Change Card Sort and Stroop Interference Task

To measure executive function, we used two tasks and created a composite score to increase reliability and validity in the measurement (Moreau et al., 2016). The scores from each task were averaged to create a unit-weighted composite score where both tasks were weighted equally. The first task was a Standard Dimensional Change Card Sort. In this task, children must sort a

set of objects two different ways: by object category and by color. First, children were given a stack of 10 cards with black and white images of fish and birds. Two boxes were placed in front of them, one marked with a picture of a black fish and the other with a picture of a white bird. The child was then asked to sort the cards by shape (fish or bird). The number of cards sorted correctly and time it took to complete the task was recorded. Next, the child was shown how to sort the cards by color with three example cards, and then was asked to sort the 10 cards by color (white or black). Again, the number of cards sorted correctly and the time it took to complete the task was recorded. For version B, the cards had white or black ships or planes. For version A, the cards were fish or bird in a 5:5 ratio, and were black or white in a 6:4 ratio. For version B, the cards were plane or ship in a 6:4 ratio and black or white in a 5:5 ratio. One participant in the 123 Ninja condition did not complete this task. A composite score of the total number of correctly sorted cards divided by the total time to sort all the cards during the second sorting was used to measure performance. The second task was a Stroop Interference task. Children were shown images of either a cat or a dog one at a time on a flashcard. In the first part of the task, children are asked to name off each image as soon as they see it, and the experimenter marks if they are correct or incorrect. Total time naming the images was also recorded. For the second part of the task, the child is asked to say the opposite animal. For example, if they see a cat, they should say dog and vice versa. Again, responses were scored as correct or incorrect based on the child's first response, and the total time naming the images was recorded. Each part of the task contained 16 images with a ratio of 1:1 for each image type. For version B children were shown images of ducks and cows. This task was adapted from the Gerstadt et al. (1994) day/night task. One participant in the ABC Ninja condition did not complete this task. A composite score of the total number of correctly named animals divided by the total time to name all the animals when the animal names were reversed was used to measure performance. The correlation coefficient between pre and post test executive function composite score was 0.69, indicating reasonable reliability.

Pearson's Picture Vocabulary Test

Vocabulary was assessed using the PPVT-4 (PPVT-4; Dunn and Dunn, 1997). A child is shown a booklet with four images on each page. The experimenter reads a word out loud and the child is asked to point to the corresponding image. The task continued until the child answered incorrectly on 10 or more words in a block. Scores were normalized with a standard score of 100. The reported standardized test-retest reliability for the PPVT is high, with a correlation coefficient of 0.91–0.94 within the age range of our participants. In our sample, the correlation coefficient between pre and post test PPVT score was 0.75, indicating reasonable reliability.

Alphabet Knowledge

Children were shown each of the 26 letters of the alphabet on a flashcard. All letters presented were uppercase letters printed in Chalkboard SE font. Two different orders were used and the order was counterbalanced across children in each condition.

Children were asked to name each letter as it was presented, and their responses were recorded. Performance was measured as the total number of correctly identified letters. The correlation coefficient between pre and post alphabet knowledge score was 0.94, indicating high reliability.

RESULTS

Training Performance

Participants in the approximate arithmetic training condition showed a consistent decrease in log difference level, indicated by the negative correlation between log difference level and trial across all training sessions ($r = -0.99$, $p < 0.0001$). The Ninja games were commercial applications and were not intended for data collection, and so measures of performance over training were less precise. At the end of each training session, the applications returned how many times the child swiped the correct symbol when it appeared. This measure indicated that across all sessions children swiped the correct numeral 63% of the time in the 123 Ninja condition, and the correct letter 67% of the time in the ABC Ninja condition. There was no evidence of a change in the number of correctly swiped letters from the first to last day of training in either Ninja condition (123 Ninja, $t = 1.26$, $p = 0.21$; ABC Ninja, $t = 0.95$, $p = 0.34$).

Analysis of Transfer Effects

Pre and post test scores for each measure are presented in Table 1. At pretest, there was no significant difference in pretest score by training condition [math composite, $F_{(2, 154)} = 0.162$, $p = 0.85$; informal math, $F_{(2, 153)} = 0.098$, $p = 0.91$; formal math $F_{(2, 153)} = 0.235$, $p = 0.79$; PPVT, $F_{(2, 153)} = 0.206$, $p = 0.81$; executive function, $F_{(2, 150)} = 0.063$, $p = 0.94$; short term memory $F_{(2, 151)} = 1.05$, $p = 0.35$; alphabet knowledge, $F_{(2, 154)} = 2.20$, $p = 0.12$; Give-a-number, Kruskal–Wallis $\chi^2 = 0.710$, $df = 2$, $p = 0.70$]. To examine change in performance from pretest to posttest gain scores were calculated for each participant for each measure. The standardized gain score for each measure was calculated by subtracting pretest score from posttest score and then dividing the gain scores by the standard deviation of the pretest scores for that measure. This allowed a comparison of gain scores across different measures. We excluded any standardized gain score when the value was smaller than $Q1 - 3 \times IQR$ or larger than $Q3 + 3 \times IQR$ (where $Q1$ is the first quartile, $Q3$ is the third quartile and IQR is in the interquartile range). This procedure removed 10 gain scores out of 1,099 data points (<1% of the data). Outlier gain scores included 1 PPVT gain score, 4 composite executive function scores, 1 informal math score, 1 formal math score, and 3 short term memory scores. Outliers were distributed across all 3 training conditions.

Transfer effects were first analyzed with an ANOVA to compare average gain score by condition for each pre/post test. This analysis collapsed across age, gender, and socioeconomic status. Contrary to our main prediction there was no significant difference in math gain score as a function of condition [Formal Math $F_{(2, 153)} = 0.956$, $p = 0.39$; Informal Math $F_{(2, 153)} = 0.133$, $p = 0.88$]. There was also no significant effect of condition on gain score for the PPVT-4 [$F_{(2, 153)} = 0.652$, $p = 0.52$],

TABLE 1 | Mean pre and post test scores (and standard deviations) for each training condition.

N	Informal math		Formal math		PPVT		Executive function composite		Short term memory		Alphabet knowledge		Number word knowledge	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Approximate arithmetic	10.673 (3.36)	11.531 (3.46)	2.735 (1.81)	2.980 (1.59)	96.837 (15.56)	100.53 (14.56)	0.104 (0.74)	0.032 (0.85)	10.837 (3.61)	11.347 (3.21)	14.143 (9.50)	14.592 (9.46)	4.347 (1.82)	4.490 (1.71)
ABC Ninja	10.755 (4.04)	11.429 (4.10)	2.653 (1.91)	2.959 (2.03)	97.429 (16.43)	100.00 (14.88)	0.081 (0.84)	0.076 (0.80)	11.592 (3.98)	12.367 (3.95)	13.204 (8.98)	15.224 (8.80)	3.939 (2.04)	4.469 (1.93)
123 Ninja	10.592 (3.52)	11.163 (3.83)	2.510 (1.95)	2.592 (1.77)	95.000 (17.38)	95.429 (17.59)	0.085 (0.76)	0.065 (0.74)	10.510 (3.31)	11.673 (3.47)	10.796 (9.87)	11.878 (9.77)	4.041 (1.98)	4.041 (2.06)

short term memory [$F_{(2, 151)} = 0.600, p = 0.55$], or executive function [$F_{(2, 150)} = 0.272, p = 0.76$]. A Kruskal–Wallis Test for nonparametric group differences revealed no effect of condition on improved knower level in the Give-a-number task (Kruskal–Wallis $\chi^2 = 4.02, df = 2, p = 0.13$). There was, however, a significant effect of condition for the alphabet knowledge test [$F_{(2, 154)} = 2.97, p = 0.05$]. Pairwise comparisons with the Holm correction indicate a significant difference between participants in the approximate arithmetic condition and ABC Ninja ($p = 0.05$), but not between 123 Ninja and ABC Ninja ($p = 0.24$) or between approximate arithmetic and 123 Ninja ($p = 0.41$). Thus, children in the ABC Ninja condition gained more knowledge of the alphabet compared to children in the approximate arithmetic condition, but not significantly more than children in the 123 Ninja condition.

We next examined whether socioeconomic status, age, math ability level, gender, experimental design factors, or training condition influenced performance on the informal and formal math test. We conducted two separate variable selection procedures to select a model that best predicted posttest informal and formal math score. We included the following variables in both variable selection procedures: pretest math composite score, training condition, training condition by pretest math composite score interaction, gender, age, whether or not the child was enrolled in NC-PreK (a proxy for SES), the version of math test the subject took at pretest (A or B), and the number of days between the pre and post test. First, stepwise model selection was performed to minimize AIC using the MASS package “stepAIC” command in R (Venables and Ripley, 2002). Both the addition and deletion of variables were allowed with this stepwise procedure. The final model selected using a minimal AIC criteria with the informal math test as the outcome included the predictors of pretest math score, approximate arithmetic condition, pretest math score by approximate arithmetic condition interaction, gender, age, and enrollment in NC-PreK (AIC = 94.64). To confirm this model, all subsets regression using the C_p statistic was conducted with the leaps package “leaps” command in R (Lumley and Miller, 2009). Using this analysis, the model derived from the minimal AIC procedure had a C_p statistic of 4.85 with 6 predictors, indicating a slight overfitting. The model that included both the main effect of 123 Ninja and pretest math score by 123 Ninja interaction as well as all the predictors from the previous model was a better fit (Table 2; $C_p = 8.82$, with 8 predictors plus the intercept). The model derived from the all subsets regression procedure with pretest math score, condition main effects, pretest score by condition interactions, age, gender, and NC-PreK enrollment as regressors is presented in Table 2.

In Table 2, all estimates are relative to the non-math control condition of ABC Ninja. Math scores were Z-scored so that estimates can be interpreted as effect sizes in terms of standard deviations. First, and most crucial to our main hypothesis, the interaction term between pretest math score and the approximate arithmetic training condition was significant [$F_{(2, 149)} = 6.48, p = 0.01$], while the main effect of the 123 Ninja condition [$F_{(2, 149)} = 0.081, p = 0.78$], and the interaction of math pretest score and 123 Ninja condition [$F_{(2, 148)} = 0.021, p = 0.88$] was

TABLE 2 | Summary of regression analyses for variables derived from model selection procedures predicting informal and formal math scores ($N = 157$).

Variable	Informal math			Formal math		
	<i>B</i>	<i>SE β</i>	<i>β</i>	<i>B</i>	<i>SE β</i>	<i>β</i>
Intercept	−1.54	0.54	−1.54**	0.955	0.08	0.067
Pretest score	0.670	0.10	0.643***	0.762	0.05	0.811***
123 Ninja condition	0.177	0.14	−0.014	−0.300	0.12	−0.168
Approximate arithmetic condition	4.39	0.14	0.111	−0.064	0.11	−0.036
Pretest score by 123 Ninja condition interaction	−0.021	0.14	−0.020	–	–	–
Pretest score by approximate arithmetic condition interaction	−0.371	0.14	−0.356*	–	–	–
Gender	−1.15	0.12	−0.307*	–	–	–
Age (in days)	0.004	0.0003	0.001***	–	–	–
NC-PreK enrollment	−1.43	0.18	−0.380*	–	–	–
R^2		0.50			0.66	
<i>F</i>		18.26***			100.5***	

All estimates are relative to the performance of the children in ABC Ninja condition. *** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$ Gender was coded with girls indicated with a 1, and boys indicated by a 0. Enrollment in NC-PreK was coded as a 1, and private school enrollment with a 0. The age variable was coded in days.

not significant. The disordinal interaction between pretest math score and the approximate arithmetic condition indicates that for children with low pretest math scores, approximate arithmetic training resulted in greater math performance at posttest than the ABC Ninja training condition. In contrast, for participants with high math scores, training with ABC Ninja resulted in better math performance.

In addition to a significant interaction of pretest math score and condition, there are also main effects of gender, age, and SES on posttest symbolic math score. On average with all else held constant, girls scored 0.307 standard deviations worse on the symbolic math post-test [$F_{(2, 149)} = 7.59, p = 0.007$] compared to boys. Age of the child was also a significant predictor of posttest symbolic math score, which is expected in a non-standardized math test. When all else was held constant a child answered 0.001 standard deviations better for every day they aged [$F_{(2, 149)} = 13.18, p = 0.0004$]. Thus on average a child answered 0.365 standard deviations better for every year of age. Finally, whether or not the child was enrolled in state funded preschool was also a significant predictor of post-test math score. On average with all else held constant, children in state funded preschools answered 0.380 standard deviations worse compared to students funded by private tuition [$F_{(2, 149)} = 6.17, p = 0.01$]. Overall, this analysis reveals that in addition to pretest math score and condition, gender, enrollment in state funded preschool, and age also impacted children's math test performance after training. Important to the goal of this experiment, accounting for the variance in informal math score due to SES, gender, and age revealed an effect of training condition for the low math scoring participants consistent with our hypothesis that approximate arithmetic training improves informal math ability.

We then ran both model selection procedures to test whether age, math ability level, SES, gender, or experimental design factors impacted performance on formal math problems. The final model selected using a minimal AIC criteria included the predictors of pretest math score and the 123 Ninja condition

(AIC = 165.47). Using the best subsets regression model selection technique, the most parsimonious model with the best C_p statistic was the same model derived from the minimal AIC criterion ($C_p = 2.72$, with 2 predictors plus the intercept). In this model, only pretest math score explained significant variance in posttest math score [$F_{(2, 154)} = 299.76, p < 0.001$] indicating that there were no effects of condition on formal math gains. The results are unchanged when the approximate arithmetic condition is added to the model, and so the model with this predictor is included in **Table 2** for better comparison of performance across all three training conditions. Contrary to our hypothesis that 123 Ninja training would improve formal math skill, this analysis indicated no effect of condition on formal math test performance.

Analysis of Transfer Effects Among Low Math Achieving Participants

Our central hypothesis was that children in the approximate arithmetic training condition would improve their informal math skill significantly more than children in the symbol identification training conditions. Contrary to this prediction, we did not find a main effect of condition among the full sample of participants. Instead, we found a significant interaction between informal math pretest score and the approximate arithmetic training condition. This interaction indicated that among participants with a low score on the informal math pretest, the approximate arithmetic training group gained more at post-test than participants in the ABC Ninja condition. Based on this finding, we reran the model in **Table 2** with data only from the participants who scored in the lower half of math pretest scores on all measures of the math test ($N = 87$)¹. Demographics of this half of the data in comparison to the full data set are shown in **Table 3**.

¹Note that participants who scored at the median pretest math score ($N = 14$) were included in the low pretest math score group.

TABLE 3 | Demographics of full sample and participants who scored in the low half of pretest math scores.

Participants	Approximate arithmetic		123 Ninja		ABC Ninja	
	All	Low math	All	Low math	All	Low math
N	53	27	52	29	52	31
Gender (males)	27	10	26	14	26	18
Age in years [mean (sd)]	4.57 (0.61)	4.37 (0.62)	4.61 (0.52)	4.48 (0.51)	4.58 (0.56)	4.41 (0.59)
Enrollment in NC-PreK	45	24	44	25	43	25

TABLE 4 | Summary of regression analyses for low pretest scoring math participants predicting informal and formal math scores ($N = 87$).

Variable	Informal math			Formal math		
	B	SE β	β	B	SE β	β
Intercept	4.56	0.69	−1.45*	0.894	0.12	−0.131
Pretest score	0.131	0.16	0.126	0.651	0.08	0.694***
123 Ninja condition	0.434	0.19	0.116	−0.217	0.16	−0.122
Approximate arithmetic condition	1.56	0.20	0.416*	0.242	0.16	0.136
Gender	−1.41	0.17	−0.375*	—	—	—
Age (in days)	0.003	0.0004	0.0008†	—	—	—
NC-PreK enrollment	−0.916	0.25	−0.244	—	—	—
R^2		0.12			0.48	
F		1.863†			25.85***	

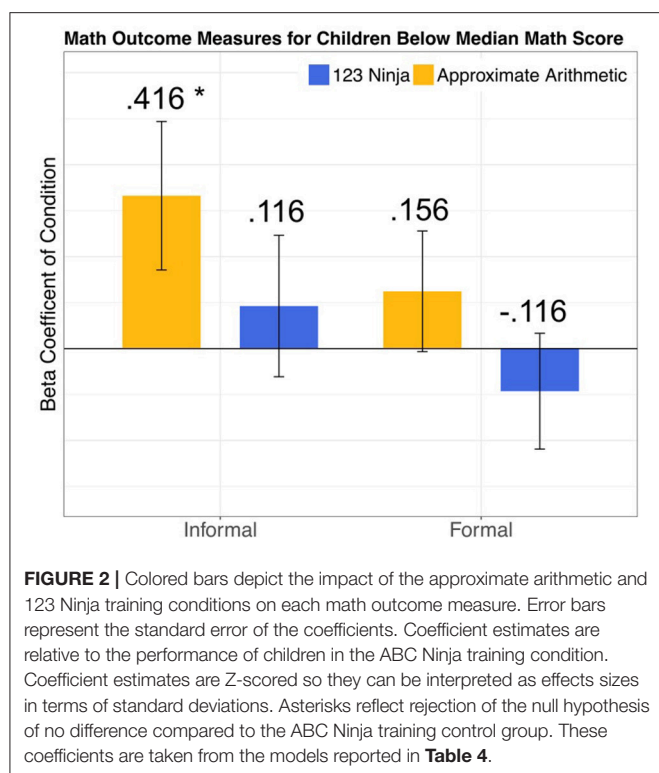
All estimates are relative to the performance of the children in ABC Ninja condition. *** $p < 0.001$, * $p < 0.05$, † $p < 0.1$. Gender was coded with girls indicated with a 1, and boys indicated by a 0. Enrollment in NC-PreK was coded as a 1, and private school enrollment with a 0. The age variable was coded in days.

Critical to our central hypothesis that approximate arithmetic training improves informal math ability, there was a significant main effect of the approximate arithmetic condition among participants with a low pretest math score [Table 4 and Figure 2; $F_{(2, 81)} = 4.24$, $p = 0.04$]. This indicates that for children with low math skills, approximate arithmetic training resulted in higher post-test informal math scores than participants in the ABC Ninja training condition. As expected, the interaction between pretest math score and condition was no longer significant among this subset of participants [$F_{(2, 80)} = 0.230$, $p = 0.81$]. The main effect of the 123 Ninja condition was also not significant [$F_{(2, 81)} = 0.363$, $p = 0.55$]. Thus, there was no effect of the 123 Ninja training condition compared to the ABC Ninja condition on post-test informal math score. Overall, these results are in line with our original hypothesis that approximate arithmetic training improves informal math skills compared to the ABC Ninja condition, however, this effect is limited to children with low initial math performance.

Among children with a low pretest math score, there was also a significant effect of gender [$F_{(2, 81)} = 4.86$, $p = 0.03$] and a marginal effect of age [$F_{(2, 81)} = 3.49$, $p = 0.07$] on informal math post-test score, but there was no longer an effect of SES [$F_{(2, 81)} = 0.966$, $p = 0.33$]. The gender effect indicates the girls scored 0.375 standard deviations worse on the symbolic math posttest compared to boys. The age effect indicates that children scored 0.0008 standard deviations better for every day they aged, or 0.292 standard deviations for every year they aged. These effects

are similar to the gender and age effects found for the full sample of participants.

Finally, similar to the results including the full sample of participants, there was no effect of condition on post-test formal math score, but there was a significant effect of pretest formal math score [Table 4; $F_{(2, 84)} = 73.178$, $p < 0.001$]. This result indicates that training condition had no impact on formal math ability. When the effects of age, gender, and enrollment in state funded preschool on formal math score are controlled, there is still no effect of condition on formal math score [Figure 2; Approximate Arithmetic $F_{(2, 81)} = 0.901$, $p = 0.35$; 123 Ninja $F_{(2, 81)} = 0.541$, $p = 0.46$]. Also consistent with the findings for the full sample of participants, children in approximate arithmetic condition with a low initial math score did not improve on measures of vocabulary [$F_{(2, 81)} = 0.000$, $p = 0.98$], short term memory [$F_{(2, 78)} = 0.001$, $p = 0.92$], executive function [$F_{(2, 81)} = 1.55$, $p = 0.22$], or number word knowledge [$F_{(2, 81)} = 0.175$, $p = 0.68$] when controlling for effects of age, gender, and state funded preschool enrollment. Children in the 123 Ninja condition also did not improve on these measures, however, they did perform significantly worse at post-test on the PPVT-4 than children in the ABC Ninja condition [$F_{(2, 81)} = 6.55$, $p = 0.01$]. Consistent with our original hypothesis, low math scoring children in the ABC Ninja condition performed significantly better on letter identification than children in both the approximate arithmetic [$F_{(2, 81)} = 7.28$, $p = 0.008$] and 123 Ninja [$F_{(2, 81)} = 5.81$, $p = 0.02$] conditions at post-test when controlling for pretest score, indicating that ABC Ninja



training improved children's letter identification skill. Overall, these results demonstrate the specificity of the approximate arithmetic training effect. Improvements in informal math skill among low math scoring approximate arithmetic participants were not due to increases in short term memory, executive function, vocabulary, or number word knowledge.

DISCUSSION

Our study was designed to ask whether approximate arithmetic training positively impacts informal, and not formal, math ability in preschool aged children over and above any benefits of two commercially available educational applications. Contrary to our hypothesis, we did not find a benefit of approximate arithmetic training on informal math performance for all participants. Instead, we found that for children with low math scores, approximate arithmetic training significantly improved informal symbolic math performance compared to training that focused on letter knowledge. While unexpected, this finding is consistent with previous research that has found the correlation between ANS acuity and math performance only among children who scored poorly on a math assessment (Bonny and Lourenco, 2013; Purpura and Logan, 2015). Consistent with our hypothesis, the positive effect of approximate arithmetic training was restricted to informal, and not formal, math abilities. We found no effect of training condition on formal math abilities, however, ABC Ninja training was effective at improving alphabet knowledge.

Previous research with our approximate arithmetic training application, Max's Math, found that approximate arithmetic

training improved the math skills of preschool children across the range of math performance (Park et al., 2016). It is important to note that the magnitude of the math standardized gain score for our approximate arithmetic training condition with all participants (0.251 with standard error 0.137) is within the standard error found for the math standardized gain score of the approximate arithmetic training condition in Park et al. (2016; 0.307 with standard error 0.070). In Park et al. (2016), the math gain score for the approximate arithmetic training group was significantly different than the math gain score for picture memory control training condition, whereas in our study among the full sample of participants the math gain scores for the symbol identification control conditions were not significantly different from the approximate arithmetic training group. It is possible that the commercially available symbol identification training games used in our study were more engaging than the picture-memory control condition used in Park et al. (2016).

A major difference between Park et al. (2016) and the current study, was that Park et al. (2016) used the TEMA-3 as an outcome measure whereas we used a modified version of the NSS. The standardized gain score for the approximate arithmetic condition in the previous study was slightly, if not significantly, higher than the gain score for the approximate arithmetic condition in the current study. The TEMA-3 may be more sensitive to the math abilities improved by approximate arithmetic training than the math measure in the current study. Additionally, the TEMA-3 is standardized to be age appropriate for children 3–9 years old, whereas the NSS was developed for children in kindergarten to 1st grade. It is possible that despite our attempt to modify the measure it was not age appropriate for preschool children.

Surprisingly, despite the fact that the 123 Ninja task was designed to teach the association between numerals and number words, children trained in this condition did not improve on our formal math test of numeral identification. In contrast, children in the ABC Ninja condition did improve in their alphabet knowledge at post-test significantly more than children in the approximate arithmetic condition. A strong possibility is that our measure of numeral identification was not as sensitive as our measure of alphabet knowledge. The numeral identification test included double-digit numbers that were not explicitly trained, while our measure of alphabet knowledge included all of the letters of the alphabet. This design resulted in a greater overlap between training and test for the ABC Ninja condition compared to the 123 Ninja condition. It is likely that with a better matched numeral identification measure 123 Ninja training would also be effective at improving numeral identification.

Another aspect of our hypothesis was that children in the approximate arithmetic condition would improve selectively on informal symbolic math skills, and not on tests of short term memory or executive function. Children in the approximate arithmetic condition with a low pretest math score improved selectively on informal math skills, and not on our pre/post test measures of executive function, short term memory, vocabulary, or number word knowledge skills. This finding suggests that improvement on informal math skills was due to the manipulation of non-symbolic quantity, and

not due to improvements on short term memory or executive function skills or to differences in number word knowledge or vocabulary.

We also found that both gender and SES influenced children's performance on the symbolic math test. Overall boys performed better on our math assessment than girls. Gender differences in performance were not reported for the NSS, the standardized math test our math measure was based upon (Research Edition: Glutting and Jordan, 2012), although work with an earlier version of the test did find a small effect of gender in the same direction as our effect (Jordan et al., 2006). We also found an effect of socio-economic status on post-test math scores consistent with previous findings (Starkey et al., 2004; Jordan et al., 2006, 2007). Indeed, Park et al. (2016) found that approximate arithmetic training was particularly effective among low income children. Our study offers more evidence that socio-economic status impacts early math learning.

In summary, consistent with our original hypothesis, approximate arithmetic training improved informal math skills significantly more than training with letter identification, however, this effect was restricted only to children with low math skill. We found a significant interaction between pretest math ability and training condition, such that low math performance participants benefitted more from approximate arithmetic training, while high math performance participants had higher post-test informal math scores after symbol identification training. Among low scoring math participants, there was a main effect of higher post-test math scores among children in the approximate arithmetic condition compared to children who trained on letter identification. As predicted, this effect was restricted to informal, and not formal, math skills. Overall, our results support the conclusion that approximate arithmetic training may be especially effective for children with low math skills, while children with a high level of math skill benefit more from symbolic training. Our study is also consistent with the general conclusion that training on educationally focused tablet applications can be effective in teaching children early academic skills.

Additional research is necessary to identify the precise conditions under which approximate arithmetic training benefits children's math learning. While we were able to demonstrate that approximate arithmetic benefits informal math ability, this category still encompasses a wide array of math skills. Future

studies should implement a larger battery of informal math questions to identify the specific math skills that benefit the most from approximate arithmetic training. Another open question is the level of math skill the child brings to the table when beginning training with approximate arithmetic. Our findings suggest that approximate arithmetic may be especially beneficial for children with low math ability. Future work should explore how math ability and factors that can broadly effect math ability, such as socio-economic status and age, interact to influence the effectiveness of intervention. Finally, our research supports the idea that approximate arithmetic training could be a useful addition to an early math curriculum, but further research is needed to understand the best way to integrate non-symbolic and approximate arithmetic into early math education. A recent convergence of work supporting the effectiveness of approximate arithmetic training, including the current study, suggests this would be a useful endeavor.

ETHICS STATEMENT

The study and protocol were reviewed and approved by the Duke University's Institutional Review Board. Written informed consent was obtained from the parents of all participants.

AUTHOR CONTRIBUTIONS

ES and EB conceived and planned the experiment. ES collected the data and performed the analyses. All authors discussed the results and contributed to the final manuscript.

FUNDING

This research was supported by a NIH R01 HD079106 award to EB.

ACKNOWLEDGMENTS

The authors would like to thank Francesca Tocci, Rachel Roberts, Primula Lane, Mary Hagan, Cayley Larimer, Lori Anne Owusu-Dapaah, Yaphet Elaias, Taylor Jones, Belex Cheng, and Chandra Swanson for help with data collection. We would also like to thank Joonkoo Park, Nick DeWind, and Stephanie Bugden for their helpful discussions of this manuscript.

REFERENCES

- Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., and Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition* 98, 199–222. doi: 10.1016/j.cognition.2004.09.011
- Barth, H., La Mont, K., Lipton, J., and Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proc. Natl. Acad. Sci. U.S.A.* 102, 14116–14121. doi: 10.1073/pnas.0505512102
- Bonny, J. W., and Lourenco, S. F. (2013). The approximate number system and its relation to early math achievement: evidence from the preschool years. *J. Exp. Child Psychol.* 114, 375–388. doi: 10.1016/j.jecp.2012.09.015
- Chen, Q., and Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychol.* 148, 163–172. doi: 10.1016/j.actpsy.2014.01.016
- Ching, B. H. H., and Nunes, T. (2017). The importance of additive reasoning in children's mathematical achievement: a longitudinal study. *J. Educ. Psychol.* 109, 477–508. doi: 10.1037/edu0000154
- Dillon, M. R., Kannan, H., Dean, J. T., Spelke, E. S., and Dufo, E. (2017). Cognitive science in the field: a preschool intervention durably enhances intuitive but not formal mathematics. *Science* 357, 47–55. doi: 10.1126/science.aal4724
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Dev. Psychol.* 43, 1428–1446. doi: 10.1037/0012-1649.43.6.1428

- Dunn, L. M., and Dunn, D. M. (1997). *Peabody Picture Vocabulary Test-Revised*. Minneapolis, MN: American Guidance Service.
- Fazio, L. K., Bailey, D. H., Thompson, C. A., and Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *J. Exp. Child Psychol.* 123, 53–72. doi: 10.1016/j.jecp.2014.01.013
- Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002
- Geary, D. C., Hoard, M. K., Nugent, L., and Bailey, D. H. (2013). Adolescent's functional numeracy is predicted by their school entry number system knowledge. *PLoS ONE* 8:e54651. doi: 10.1371/journal.pone.0054651
- Gerardi, K., Goette, L., and Meier, S. (2013). Numerical ability predicts mortgage default. *Proc. Natl. Acad. Sci. U.S.A.* 110, 11267–11271. doi: 10.1073/pnas.1220568110
- Gerstadt, C. L., Hong, Y. J., and Diamond, A. (1994). The relationship between cognition and action: performance of children 3 1/2 - 7 years old on a Stroop like day-night task. *Cognition* 53, 129–153. doi: 10.1016/0010-0277(94)90068-X
- Ginsburg, H. P., and Baroody, A. J. (2003). *Test of Early Mathematics Ability (3rd edn)*. Austin, TX: Pro-Ed.
- Glutting, J., and Jordan, N. C. (2012). *Number Sense Screener*. Baltimore, MD: Brookes Publishing.
- Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., and Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proc. Natl. Acad. Sci. U.S.A.* 109, 11116–11120. doi: 10.1073/pnas.1200196109
- Halberda, J., Mazocco, M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nat. Lett.* 455, 665–668. doi: 10.1038/nature07246
- Hirsh-Pasek, K., Zosh, J. M., Golinkoff, R. M., Gray, J. H., Robb, M. B., and Kaufman, J. (2015). Putting education in “educational” apps lessons from the science of learning. *Psychol. Sci. Public Interest* 16, 3–34. doi: 10.1177/1529100615569721
- Hyde, D. C., Khanum, S., and Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition* 131, 92–107. doi: 10.1016/j.cognition.2013.12.007
- Jordan, N. C., Glutting, J., and Ramineni, C. (2010). The importance of number sense to mathematics achievement in first and third grades. *Learn. Individ. Diff.* 20, 82–88. doi: 10.1016/j.lindif.2009.07.004
- Jordan, N. C., Kaplan, D., Locuniak, M. N., and Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learn. Disabil. Res. Pract.* 22, 36–46. doi: 10.1111/j.1540-5826.2007.00229.x
- Jordan, N. C., Kaplan, D., Nabors Oláh, L., and Locuniak, M. N. (2006). Number sense growth in kindergarten: a longitudinal investigation of children at risk for mathematics difficulties. *Child Dev.* 77, 153–175. doi: 10.1111/j.1467-8624.2006.00862.x
- Jordan, N. C., Kaplan, D., Ramineni, C., and Locuniak, M. N. (2009). Early math matters: kindergarten number competence and later mathematics outcomes. *Dev. Psychol.* 45, 850–867. doi: 10.1037/a0014939
- Khanum, S., Hanif, R., Spelke, E. S., Berteletti, I., and Hyde, D. C. (2016). Effects of non-symbolic approximate number practice on symbolic numerical abilities in Pakistani children. *PLoS ONE* 11:e0164436. doi: 10.1371/journal.pone.0164436
- Kibbe, M. M., and Feigenson, L. (2015). Young children “solve for x” using the approximate number system. *Dev. Sci.* 18, 38–49. doi: 10.1111/desc.12177
- Libertus, M. E., Feigenson, L., and Halberda, J. (2013). Numerical approximation abilities correlate with and predict informal but not formal mathematics abilities. *J. Exp. Child Psychol.* 116, 829–838. doi: 10.1016/j.jecp.2013.08.003
- Lumley, T., and Miller, A. (2009). *Leaps: Regression Subset Selection*. R package version 3.0. Available online at: <http://CRAN.R-project.org/package=leaps>
- McCrink, K., and Spelke, E. S. (2010). Core multiplication in childhood. *Cognition* 116, 204–216. doi: 10.1016/j.cognition.2010.05.003
- McCrink, K., and Spelke, E. S. (2016). Non-symbolic division in childhood. *J. Exp. Child Psychol.* 142, 66–82. doi: 10.1016/j.jecp.2015.09.015
- Moreau, D., Kirk, I. J., and Waldie, K. E. (2016). Seven pervasive statistical flaws in cognitive training interventions. *Front. Hum. Neurosci.* 10:153. doi: 10.3389/fnhum.2016.00153
- Nunes, T., Bryant, P., Evans, D., Bell, D., Gardner, S., Gardner, A., et al. (2007). The contribution of logical reasoning to the learning of mathematics in primary school. *Br. J. Dev. Psychol.* 25, 147–166. doi: 10.1348/026151006X153127
- Park, J., Bermudez, V., Roberts, R. C., and Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *J. Exp. Child Psychol.* 152, 278–293. doi: 10.1016/j.jecp.2016.07.011
- Park, J., and Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychol. Sci.* 24, 2013–2019. doi: 10.1177/0956797613482944
- Park, J., and Brannon, E. M. (2014). Improving arithmetic performance with number sense training: an investigation of underlying mechanism. *Cognition* 133, 188–200. doi: 10.1016/j.cognition.2014.06.011
- Purpura, D. J., and Logan, J. A. (2015). The nonlinear relations of the approximate number system and mathematical language to early mathematics development. *Dev. Psychol.* 51, 1717–1724. doi: 10.1037/dev0000055
- Reyna, V. F., Nelson, W. L., Han, P. K., and Dieckmann, N. F. (2009). How numeracy influences risk comprehension and medical decision making. *Psychol. Bull.* 135, 943–973. doi: 10.1037/a0017327
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372
- Starkey, P., Klein, A., and Wakeley, A. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Child. Res. Q.* 19, 99–120. doi: 10.1016/j.ecresq.2004.01.002
- Szucs, D., and Myers, T. (2017). A critical analysis of design, facts, bias and inference in the approximate number system training literature: a systematic review. *Trends Neurosci. Educ.* 6, 187–203. doi: 10.1016/j.tine.2016.11.002
- Venables, W. N., and Ripley, B. D. (2002). *Modern Applied Statistics with S (Statistics and Computing)*, 4th Edn. New York, NY: Springer.
- Wynn, K. (1990). Children's understanding of counting. *Cognition* 36, 155–193.
- Wynn, K. (1992). Children's acquisition of the number words and the counting system. *Cogn. Psychol.* 24, 220–251.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Szkudlarek and Brannon. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Testing the Efficacy of Training Basic Numerical Cognition and Transfer Effects to Improvement in Children's Math Ability

Narae Kim¹, Selim Jang^{1,2} and Soohyun Cho^{1*}

¹ Department of Psychology, Chung-Ang University, Seoul, South Korea, ² Department of Psychology, University of Illinois at Urbana-Champaign, Champaign, IL, United States

OPEN ACCESS

Edited by:

Marcus Lindskog,
Uppsala University, Sweden

Reviewed by:

Paula Goolkasian,
University of North Carolina
at Charlotte, United States
Maria Grazia Di Bono,
Università degli Studi di Padova, Italy

*Correspondence:

Soohyun Cho
soohyun@cau.ac.kr

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 15 May 2018

Accepted: 03 September 2018

Published: 02 October 2018

Citation:

Kim N, Jang S and Cho S (2018)
Testing the Efficacy of Training Basic
Numerical Cognition and Transfer
Effects to Improvement in Children's
Math Ability. *Front. Psychol.* 9:1775.
doi: 10.3389/fpsyg.2018.01775

The goals of the present study were to test whether (and which) basic numerical abilities can be improved with training and whether training effects transfer to improvement in children's math achievement. The literature is mixed with evidence that does or does not substantiate the efficacy of training basic numerical ability. In the present study, we developed a child-friendly software named "123 Bakery" which includes four training modules; non-symbolic numerosity comparison, non-symbolic numerosity estimation, approximate arithmetic, and symbol-to-numerosity mapping. Fifty-six first graders were randomly assigned to either the training or control group. The training group participated in 6 weeks of training (5 times a week, 30 minutes per day). All participants underwent pre- and post-training assessment of their basic numerical processing ability (including numerosity discrimination acuity, symbolic/non-symbolic magnitude estimation, approximate arithmetic, and symbol-to-numerosity mapping), overall math achievement and intelligence, 6 weeks apart. The acuity for numerosity discrimination (approximate number sense acuity; hereafter ANS acuity) significantly improved after training, but this training effect did not transfer to improvement in symbolic, exact calculation, or any other math ability. We conclude that basic numerical cognition training leads to improvement in ANS acuity, but whether this effect transfers to symbolic math ability remains to be further tested.

Keywords: approximate number sense, training, numerosity comparison, numberline estimation, approximate arithmetic, symbol-to-numerosity mapping

INTRODUCTION

The ability to process numerosity information is essential for everyday life in both humans and animals (Agrillo et al., 2008; Libertus et al., 2011; Leibovich et al., 2017). Approximate number sense (ANS) enables the ability to grasp approximately how many items there are and to roughly add or subtract sets of items. Some researchers believe that basic numerical processing ability and higher level mathematical achievement builds on the ANS (Rousselle and Noël, 2007; De Smedt and Gilmore, 2011; Sasanguie et al., 2012, 2013; Jang and Cho, 2018). Basic numerical processing includes numerosity comparison, symbolic number comparison, numberline estimation, and understanding the mapping between symbolic numbers and their corresponding numerosity (or non-symbolic magnitude), etc. Basic numerical processing abilities are reported to predict future

math achievement (Jordan et al., 2007; Locuniak and Jordan, 2008; Lyons and Beilock, 2011; Mazzocco et al., 2011; Sasanguie et al., 2013; Starr et al., 2013; Martin et al., 2014). Furthermore, children with mathematical learning disabilities or developmental dyscalculia have been found to show low performance on basic numerical processing (Rousselle and Noël, 2007; Geary et al., 2008; De Smedt et al., 2009; Piazza et al., 2010; De Smedt and Gilmore, 2011). Some studies reported that training on basic numerical abilities led to improvement in math achievement (Park and Brannon, 2013, 2014; Park et al., 2016; Sella et al., 2016). However, different types of training were used across studies and the reports of the efficacy of training were mixed. Thus, at present it is not easy to draw a conclusion on whether or not intervention on basic numerical abilities can improve one's math performance (Schneider et al., 2016; Szűcs and Myers, 2017).

In some studies, training on approximate arithmetic (approximate addition and subtraction) using dot arrays improved the training groups' symbolic addition/subtraction abilities compared to the control group (Park and Brannon, 2013, 2014; Hyde et al., 2014; Khanum et al., 2016; Park et al., 2016; Au et al., 2018; Szkudlarek and Brannon, 2018). In contrast, Räsänen et al. (2009) did not find any improvement on arithmetic (addition and subtraction) and counting abilities after training with the Number Race program¹ (Wilson et al., 2006) although children's ANS acuity was improved (Räsänen et al., 2009). Szkudlarek and Brannon (2018) reported that approximate arithmetic vs. numeral identification training was effective for preschoolers with low vs. high math skills, respectively. But the training improved only early informal, but not formal, math skills. Based on a meta analysis, Szűcs and Myers (2017) concluded that, presently, there is no evidence that ANS training improves symbolic arithmetic given methodological issues and heterogeneity across studies. One crucial issue relates to the inclusion of symbolic arithmetic practice within the training program itself (Wilson et al., 2006, 2009; Räsänen et al., 2009; Vilette et al., 2010; Kucian et al., 2011; Obersteiner et al., 2013; Sella et al., 2016). In these cases, improvement in symbolic math ability after repeated practice of symbolic arithmetic may simply reflect practice (or test-retest) effect rather than a true transfer effect of training. Furthermore, many studies reporting a significant training effect tended to have small effect sizes or unstable results (e.g., being influenced by outliers) (Szűcs and Myers, 2017).

The goal of the present study was to investigate whether (and which) basic numerical abilities can be improved with training and to test whether the training effect transfers to improvement in overall math achievement. We developed a child-friendly computer based software named "123 Bakery" which included four modules for training basic numerical abilities (numerosity comparison, numberline estimation, approximate, non-symbolic addition/subtraction, and symbol-to-numerosity mapping). Exact, symbolic arithmetic practice during training was

purposefully excluded in order to thoroughly test whether the effect of training on basic numerical cognition truly transfers to exact, symbolic math ability without explicit practice in this domain. Our software was designed to include several training modules within each session, as in typical educational interventions (Kroesbergen and Van Luit, 2003; Gersten et al., 2009; Coddington et al., 2011). Training sessions were administered at the child's home which increased ecological validity of our training to real-world educational applications. All assessments were administered at the child's home as well. The difficulty level of training was tailored to each participant to help participants learn in their own zone of proximal development (i.e., adaptive training). Training effects were tested by comparing assessment scores acquired immediately before and after training. In order to measure improvement of trained abilities while minimizing test-retest effects, we designed tasks with alternative visual interfaces (see Materials and Methods for details). Mathematical achievement was assessed with a comprehensive standardized math test battery (which included number concept, arithmetic, geometry, and problem solving) and a computerized arithmetic test. We induced intensive home training over 6 weeks (5 days/week, 35 min/day) which is by far the longest in the total duration of training compared to previous studies. The range of numerosity was also much extended (up to 300 depending on performance) which was much larger than most previous studies (which included numerosities up to 80) (Wilson et al., 2006, 2009; Räsänen et al., 2009; Kucian et al., 2011; Park and Brannon, 2013, 2014; Park and Brannon, 2014; Hyde et al., 2014; Park et al., 2016; Sella et al., 2016; Au et al., 2018). In other words, the present study aimed to rigorously test the efficacy of training basic numerical abilities based on sufficiently long durations across a large range of magnitudes while minimizing the influence of test-retest effects. The duration of training in the present study was longer than that of other lab-based training studies (but it was similar to the average duration of intervention/training programs commonly used in real-world educational settings (Cohen et al., 1982; Kroesbergen and Van Luit, 2003). Furthermore, we carefully controlled for non-numerical visual properties of the non-symbolic stimuli (dot arrays) during training and assessment, so that the influence of non-numerical visual magnitudes can be minimized. Finally, our home-based training procedure improved the ecological validity of our training program enabling more confident generalization to real-world, educational applications compared to studies which conducted training in lab settings (Hyde et al., 2014).

Given inconsistencies in the literature, we did not have an *a priori* hypothesis in favor of the idea that basic math abilities can be improved with training and that such training effects will be transferred to improvement in math achievement (especially when exact calculation is not included in the training). By using sufficiently long duration of training and wide range of magnitudes (while controlling for the influence of non-numerical visual magnitudes), we did not make type II error due to insufficient duration/range of training or contamination by extraneous variables.

¹ The Number Race program includes numerosity comparison, mapping between symbolic and non-symbolic number, and symbolic addition/subtraction.

MATERIALS AND METHODS

Participants

Fifty-six 1st graders participated in the study. Data from 10 children who did not complete the experiment or whose performance (on pre-training numerosity comparison or the final level reached on the training modules) was lower than 2 SDs below the mean were excluded. (See **Supplementary Materials** for further details.) Thus, data from forty-six children were included in the analysis (24 females; mean age = 7.70 years; and $SD = 0.30$). Participants were recruited

by advertisement. All participants and their parents provided written informed consent before participation. The IRB committee of Chung-Ang University approved all protocols of the study (IRB-2013-55). Participants were randomly assigned to either the training or control group. Participants received monetary compensation after completion of the experiment.

Procedure

Participants were randomly assigned to either the training ($n = 22$) or control ($n = 24$) group. Pre-training assessments

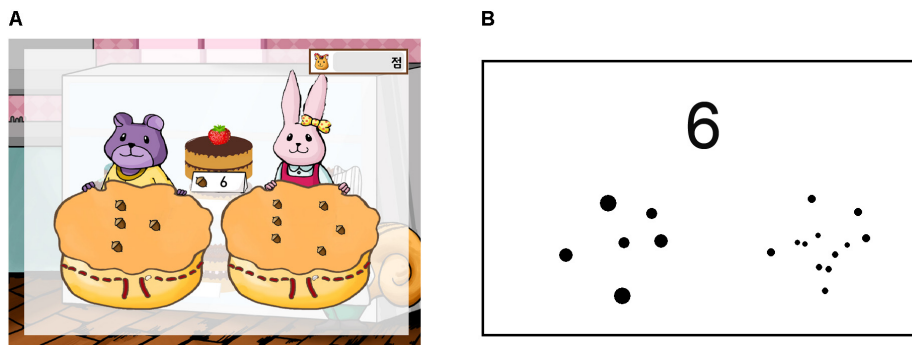


FIGURE 1 | Screenshots of an example trial of (A) the “Selling Cakes” module of “123 Bakery” and (B) “Symbol-to-Numerosity Mapping” task used for pre- and post-training assessment.

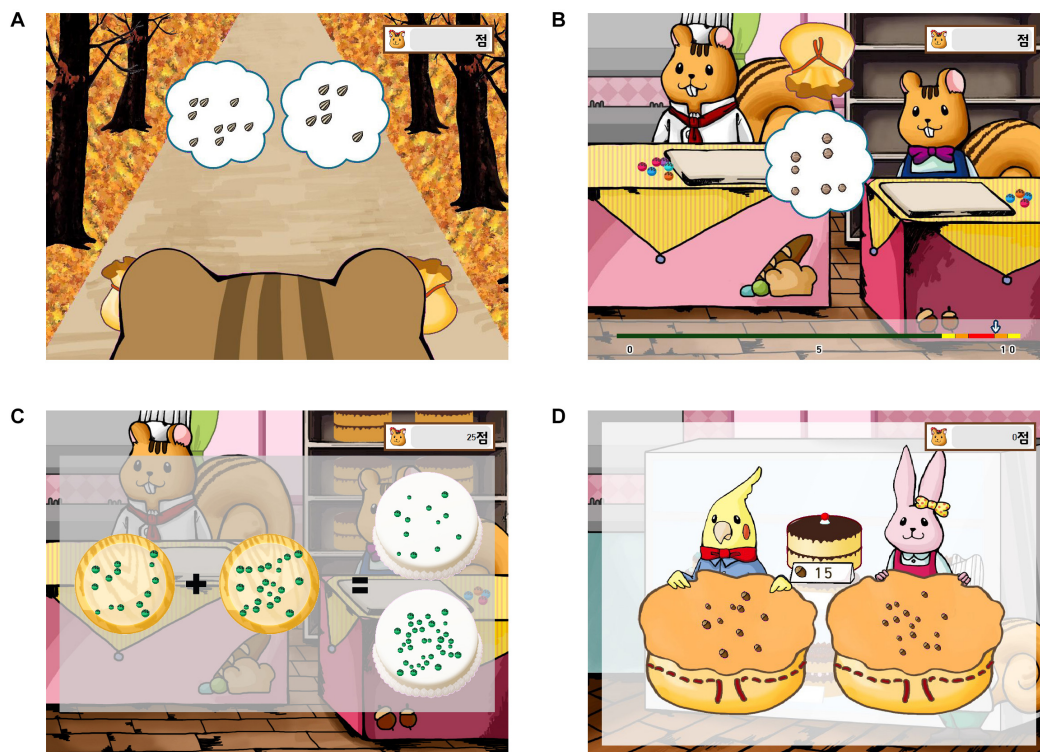


FIGURE 2 | Screenshots of example trials from (A) “Gathering Ingredients,” (B) “Guess How Many?,” (C) “Cake Decoration,” and (D) “Selling Cakes” modules of “123 Bakery.”



FIGURE 3 | Example trial of the numerosity comparison task.

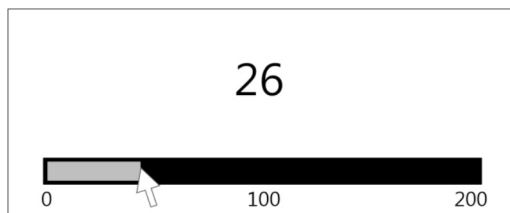


FIGURE 4 | Example trial of the symbolic numberline estimation task with an endpoint of 200.

[including basic numerical processing tasks, two math achievement tests, and the Raven's Advanced Progressive Matrices (APMs) test] were administered to all participants. Only the training group participated in 30 training sessions over 6 weeks using a computerized software ("123 Bakery"). After 6 weeks had passed since the administration of the pre-training assessment, all participants were administered post-training assessment. The pre- and post-training assessments of basic numerical abilities were conducted with four tasks corresponding to each training module using alternate visual

formats in order to minimize practice or test–retest effects at the visuomotor level (see **Figure 1**).

Materials

Basic Numerical Cognition Training Program "123 Bakery"

We developed a computerized program named "123 Bakery" which composed of four training modules. The four training modules included (1) numerosity comparison ("Gathering Ingredients"), (2) non-symbolic numberline estimation ("Guess How Many?"), (3) Approximate Addition & Subtraction ("Cake Decoration"), and (4) Symbol-to-Numerosity Mapping ("Selling Cakes"). (Each training module is explained in the next section.) Each module was 6 minutes long. Feedback on the correctness of the response was provided after each trial. The cumulative total score (within each session) was updated real-time and was always shown on the top right-hand side of the screen (**Figure 2**). Task difficulty increased as subjects mastered each Level by accomplishing a certain degree of performance accuracy (0.7–0.9 accuracy among the last 10–20 trials depending on the Level; see **Supplementary Tables S1–S4** for details).

In order to control for the influence of non-numerical visual properties of dot arrays (e.g., individual dot size, cumulative surface area, and convex hull) during numerosity processing, we made convex hull equivalent for all dot arrays and divided trials into two control conditions (area vs. size controlled conditions) (Pica et al., 2004; Halberda and Feigenson, 2008; Jang and Cho, 2016; Park and Cho, 2016; Lee and Cho, 2017). First, on half of the trials, dot arrays were matched on cumulative surface area (area controlled condition) and on the other half of the trials, dot arrays were matched on individual dot size (size controlled condition). The order of trial presentation was randomly intermixed.

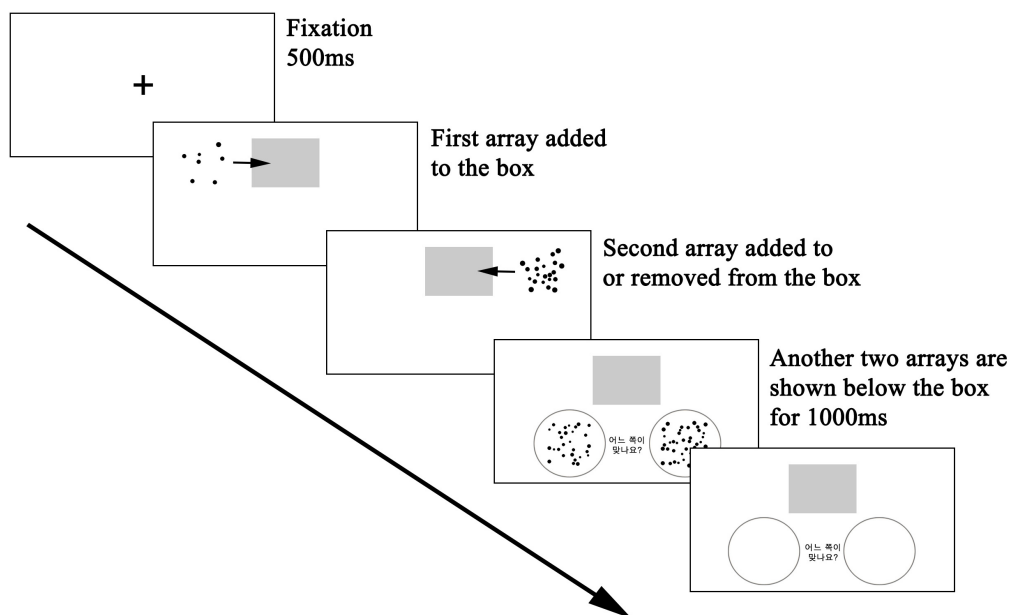


FIGURE 5 | Example trial procedure of the approximate arithmetic task.

Although it is not possible to perfectly control for the influence of non-numerical visual properties of dot arrays during numerosity processing, the use of randomly intermixed control conditions and making convex hull equivalent across all dot arrays ensured that non-numerical visual magnitude could not be reliably used as an alternative cue to guess numerosity (Maloney et al., 2010; Gebuis and Reynvoet, 2012a,b; Leibovich and Henik, 2013; Dietrich et al., 2015).

Training module 1: numerosity comparison (“gathering ingredients”)

Two arrays (of berries or nuts) appeared side by side for 1,000 ms. Subjects were instructed to choose the more numerous array (Figure 2A). Task difficulty increased as the set size became larger (range = 6–200) and as the ratio of magnitudes approached 1 (range = 2:3–9:10). Audiovisual feedback on the correctness of the response was provided after each trial.

Training module 2: non-symbolic Numberline Estimation (“Guess How Many?”)

Subjects were presented with an array (of berries or nuts, etc.) for 1,000 ms at the center of the screen. Subjects were asked to click on a location on the numberline which corresponds to the estimated numerosity of the elements of the array (Figure 2B). If the estimate was within the “accurate zone” (see Supplementary Table S2 for details), positive feedback was given. Task difficulty varied by the numerosity of the stimulus, the maximum value (end point) of the numberline, and the relative width (i.e., proportion) of the accurate zone.

Training module 3: non-symbolic addition/subtraction (“cake decoration”)

Subjects were presented with two arrays (of berries or nuts, etc.) for 1,000 ms and were asked to perform approximate addition or subtraction. Then, two arrays were additionally shown as options to choose from. Subjects were asked to respond by choosing one of the two options which seemed closer to their approximate answer within 6 s (Figure 2C). Audiovisual feedback on the correctness of the response was provided after each trial.

Training module 4: symbol-to-numerosity mapping (“selling cakes”)

Subjects were asked to choose which animal character (the customer) possessed the correct number of nuts which corresponded to the price of the cake (shown as a numeral at the center of the screen) (Figure 2D). Task difficulty increased as the ratio of numerosities approached 1 and as the price of the cake increased. Audiovisual feedback on the correctness of the response was provided after each trial.

Pre- and Post-training Assessments

Four basic numerical processing tasks

The four basic numerical processing tasks (numerosity comparison, symbolic and non-symbolic numberline estimation, approximate arithmetic, and symbol-to-numerosity mapping) had the same structure as the four training modules except that

TABLE 1 | Descriptive statistics of performance from the pre- and post-training assessments of basic numerical processing abilities.

	Pre-training			Post-training		
	Training	Control	t-test	Training	Control	t-test
Numerosity comparison	ACC RT PAE	0.75 (0.07) 1, 532.02 (315.04) 0.05 (0.02)	1.30 ($p = 0.20$) -42 ($p = 0.68$) -0.22 ($p = 0.82$)	0.81 (0.07) 1, 460.85 (345.24) 0.05 (0.01)	0.71 (0.08) 1, 428.15 (304.51) 0.05 (0.02)	4.37 ($p < 0.001$) 0.34 ($p = 0.73$) -1.51 ($p = 0.14$)
Symbolic numberline estimation	RT PAE	3, 706.75 (1, 316.72) 0.14 (0.03)	-0.30 ($p = 0.77$) -0.22 ($p = 0.83$)	2, 710.71 (623.46) 0.16 (0.05)	3, 576.84 (993.24) 0.14 (0.03)	-3.50 ($p < 0.001$) 1.78 ($p = 0.08$)
Non-symbolic numberline estimation	RT	2, 549.57 (881.61)	-0.35 ($p = 0.73$)	2, 664.78 (814.05)	2, 772.19 (1, 023.36)	-0.39 ($p = 0.70$)
Non-symbolic addition	ACC RT	0.60 (0.10) 1, 959.34 (763.18)	1.38 ($p = 0.18$) -0.76 ($p = 0.45$)	0.59 (0.08) 2, 053.98 (938.94)	0.59 (0.08) 2, 028.04 (613.66)	-0.10 ($p = 0.93$) 0.11 ($p = 0.91$)
Non-symbolic subtraction	ACC RT	0.53 (0.08) 1, 933.03 (619.19)	1.11 ($p = 0.28$) -1.32 ($p = 0.19$)	0.55 (0.06) 2, 305.66 (949.68)	0.54 (0.09) 2, 476.66 (2, 220.41)	0.47 ($p = 0.64$) -0.33 ($p = 0.74$)
Symbol-to-numerosity mapping	ACC RT	0.60 (0.06) 1, 120.88 (290.35)	0.23 ($p = 0.82$) -1.76 ($p = 0.09$)	0.61 (0.06) 1, 165.01 (304.71)	0.58 (0.07) 1, 339.36 (0.377.88)	1.33 ($p = 0.19$) -1.71 ($p = 0.09$)
Raven's APM		0.43 (0.19)	0.81 ($p = 0.43$)	0.55 (0.19)	0.48 (0.19)	1.227 ($p = 0.23$)

Numbers in parentheses denote SDs. The results of between-group t-tests at each time (pre-training and post-training) for each assessment scores are also shown.

the array consisted of black dots on a white background. Each task is explained in the following sections.

Numerosity comparison. Subjects were presented with a pair of dot arrays (1,000 ms) and were asked to choose the array with the greater number of dots (**Figure 3**). Subjects pressed the #3 key for the array on the left and #8 key for the array on the right. The left–right location of the correct answer was counterbalanced. The ratio of numerosities included 1:2, 3:4, 5:6, 6:7, 7:, and 8:9. The entire stimulus list is shown in **Supplementary Table S5**. There was a total of 120 trials.

Symbolic and non-symbolic numberline estimation. The numberline estimation task was conducted using both symbolic (Arabic numerals) and non-symbolic magnitudes (**Figure 4**). The trials were divided into two blocks based on the value of the end point of the numberline (100 or 200). The stimuli included 5, 18, 32, 55, 73, and 98 for block 1 and 5, 18, 42, 78, 111, 133, 147, 172, and 187 for block 2. The target stimulus appeared for 1,000 ms.

The accuracy of performance was calculated with Percent Absolute Error (PAE; Eq. 1) (Siegler and Booth, 2004). Smaller PAE represents smaller error in estimation and greater linearity in mental magnitude representations (Siegler and Booth, 2004; Booth and Siegler, 2006). For each target stimulus, three trials were repeated. The mean PAE for each target stimulus was used as the dependent variable (Siegler and Ramani, 2009).

$$\text{PAE} = \frac{\text{Estimate} - \text{Estimated Magnitude}}{\text{The scale of the Numberline}}$$

Eq. 1. Calculation of PAE

Approximate arithmetic. The procedure of the task was similar to the Approximate Arithmetic condition used in Park and Brannon (2013, 2014). This task was administered in two separate blocks for addition and subtraction. Subjects were first shown a dot array which was added to a gray box (**Figure 5**). Next, another dot array was either added to or removed from the gray box. Finally, the subjects chose one of two new dot arrays whose numerosity seemed closer to the perceived total number of dots in the gray box. They responded by pressing the #3 key to choose the array on the left and #8 key for the array on the right. Task difficulty was manipulated by the ratio of the set sizes of the two arrays (4:5, 4:6, and 4:7) presented as a pair on each trial. Addition and subtraction were performed on arrays with numerosities ranging from 6–51.

The numerosity of arrays presented as options ranged from 16–91. Including 5 practice trials, a total of 35 trials were administered.

Symbol-to-numerosity mapping. Subjects were asked to choose one of two arrays presented for 1000 ms whose numerosity matched the Arabic number presented at the center of the screen (**Figure 1B**). The ratio between the magnitude of the stimuli varied from 1:1.75 to 4:5 (1:1.75, 1:2, 2:3, 3.5:5, 3:4, and 4:5). The set size of the stimuli varied from 6 to 100 (6–30, 30–50, and 50–100). The left–right position of the correct answer was counterbalanced. A total of 144 trials were administered. The order of trials from each ratio/condition was randomly intermixed.

Mathematical Achievement Tests

Comprehensive math achievement test (KNISE-BAAT)

The Korean National Intelligence for Special Education–Basic Academic Achievement Test (KNISE-BAAT for math) (Park et al., 2008) was used to measure mathematical performance. KNISE-BAAT consists of four subdomains (number concept, arithmetic, geometry, and problem solving).

Computerized arithmetic task

Subjects solved 64 problems of addition and subtraction without paper and pencil on a computer. Three ranges of numbers were used (6–30, 30–50, and 50–99). Participants were instructed to type the answer using the number keys on the keyboard. There was no time limit. (Accuracy rather than RT of problem solving was emphasized). Thus, accuracy rather than RT was the main dependent variable of interest.

Raven's APM test

Children's fluid intelligence was measured with an abbreviated version of the Raven's APM test (Arthur et al., 1999). This score was used as a covariate in order to control individual differences in fluid intelligence.

RESULTS

Test of Between-Group Differences in Pre/Post-training Assessments

The training and control group were matched on age and gender. Independent samples *t*-tests revealed no difference in age [$t(44) = 1.99, p = 0.05$] and gender [$t(44) = 0.86, p = 0.39$] between groups. In addition, our groups did not differ on

TABLE 2 | The result of mixed 2×2 repeated measures ANOVA on numerosity comparison accuracy with group as the between-subject factor and time as the within-subject factor.

Dependent variable		Source	SS	df	MS	F	P	η^2
Numerosity comparison accuracy	Within subjects	Time	0.02	1	0.02	4.05	0.05	0.08
		Group \times time	0.03	1	0.03	7.47	0.01	0.15
		Error	0.16	44	0.01			
	Between subjects	Group	0.10	1	0.10	11.01	0.01	0.20
		Error	0.40	44	0.01			

TABLE 3 | Pre- and post-training assessment of math achievement (computerized arithmetic, KNISE-BAAT) of the training and control groups.

	Pre-training			Post-training		
	Training	Control	t-test	Training	Control	t-test
Computerized arithmetic	0.89 (0.13)	0.88 (0.17)	0.322 ($p = 0.75$)	0.92 (0.06)	0.86 (0.15)	1.86 ($p = 0.07$)
Reaction time	4,505.18 (1,174.94)	6,418.37 (5,249.26)	-1.67 ($p = 0.10$)	4,158.39 (1,673.01)	5,442.65 (3,251.20)	-1.66 ($p = 0.10$)
KNISE-BAAT	13.50 (2.39)	12.54 (2.47)	1.34 ($p = 0.19$)	15.32 (2.64)	14.33 (3.07)	1.16 ($p = 0.25$)
Number concept	13.55 (3.50)	13.88 (3.27)	-0.33 ($p = 0.74$)	14.41 (2.63)	14.96 (2.96)	-0.66 ($p = 0.51$)
Geometry	15.68 (3.03)	14.29 (3.54)	1.42 ($p = 0.16$)	17.23 (2.84)	15.92 (2.62)	1.63 ($p = 0.11$)
Arithmetic	14.68 (4.34)	14.04 (4.60)	0.48 ($p = 0.63$)	17.23 (3.58)	16.79 (3.76)	0.40 ($p = 0.69$)
Problem solving	57.41 (9.45)	54.75 (11.31)	0.86 ($p = 0.39$)	64.18 (9.02)	61.92 (10.18)	0.80 ($p = 0.43$)
Total score						

Numbers in parentheses denote SDs. The results of between-group *t* tests at each time (pre-training, post-training) for each assessment score are also shown.

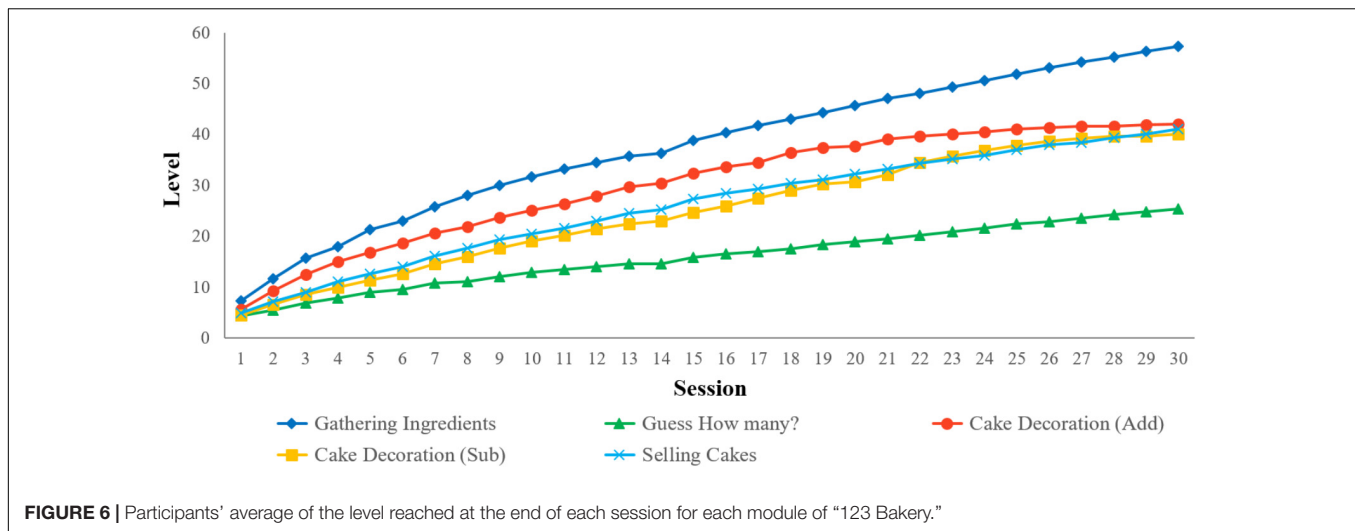
pre- and post-training assessments of math achievement or fluid intelligence ($p > 0.05$; **Table 1**).

Training Effects on Basic Number Processing Abilities

Descriptive statistics of performance from pre- and post- training assessments of basic number processing abilities are provided in **Table 1**. The training group's performance at the end of each session for each module of "123 Bakery" are shown in **Figure 6**. The final Level reached at the end of training for each module of "123 Bakery" and the overall average of the 30 mean performance scores (accuracy and RT) from each session are provided in **Supplementary Table S6**. In order to test for training effects, a 2×2 mixed repeated measures ANOVA was conducted on basic number processing performance (numerosity comparison, symbolic and non-symbolic numberline estimation, symbol-to-numerosity mapping, and approximate arithmetic) with time (pre-, post-training) as the within-subject factor and group (training, control) as the between-subject factor (**Table 2**). A significant two-way interaction would indicate the presence of a training effect that is selective for the training group compared to the control group. The two-way interaction between time and group was significant only for numerosity comparison accuracy [$F(1,44) = 7.47$; $p < 0.01$, *partial* $\eta^2 = 0.15$, **Figure 7, Table 2**]. (No other interaction effects were significant ($p > 0.05$; see **Supplementary Table S7** for results of the mixed repeated measures ANOVAs on other measures of basic number processing abilities). Given the significant two-way interaction effect, *post hoc* tests of the simple main effects of group at each time (pre-training, post-training) were conducted for numerosity comparison accuracy. A significant training effect should be manifested as higher post-training (but not pre-training) performance of the training group compared to the control group. There was no difference in numerosity comparison accuracy between the training and control group at pre-training, but the training group had significantly higher numerosity comparison accuracy at post-training [pre-training: $t(44) = 1.30$, $p = 0.20$, post-training: $t(44) = 4.37$, $p < 0.001$, **Figure 7, Table 1**]. In other words, training with the "Gathering Ingredients" module improved the training group's numerosity comparison accuracy.

Transfer Effects to Math Achievement

Descriptive statistics of performance from pre- and post-training assessments of math achievement are provided in **Table 3**. The results of independent samples *t*-tests at each time (pre-training, post-training) for each assessment score are also shown in **Table 3**. In order to investigate whether the effect of training transfers to improvement on math achievement, a mixed repeated measures ANOVA was conducted on all mathematical achievement scores (KNISE-BAAT and computerized arithmetic) with time (pre-, post-training) as the within-subject factor and group (training, control) as the between-subject factor (**Table 4**). There were no significant interaction effects on either KNISE BAAT or computerized arithmetic scores ($p > 0.05$, **Table 4**; see

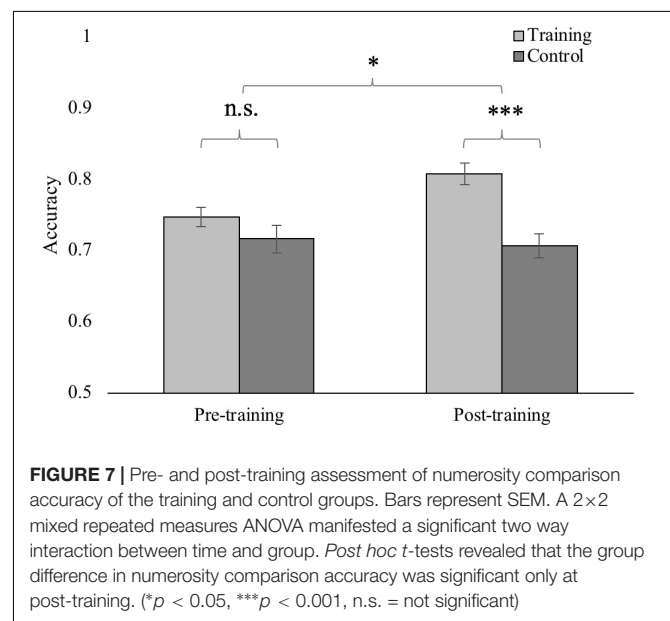


Supplementary Table S7 for results of the mixed repeated measures ANOVA on all other measures of math achievement).

DISCUSSION

The present study examined whether or which basic numerical processing ability can be improved with training and whether this training effect can be transferred to improvement in different domains of mathematical achievement. We developed a child-friendly training program called "123 Bakery" which included four training modules ("Gathering Ingredients," "Guess How Many?," "Cake Decoration," and "Selling Cakes"). The dot arrays used as stimuli representing non-symbolic magnitude were controlled so that the influence of non-numerical visual properties was minimized. Exact, symbolic calculation was purposefully excluded from training in order to examine whether training on basic numerical ability improves exact, symbolic calculation while ruling out direct practice or test-retest effects. All participants were assessed on their basic numerical ability twice, 6 weeks apart. The training group participated in 6 weeks of training immediately after the first assessment. The second assessment took place immediately after the training session ended. Compared to the control group, the numerosity comparison accuracy of the training group improved significantly more at post-training assessment. This result is consistent with previous studies reporting improvement of ANS acuity after training (DeWind and Brannon, 2012; Odic et al., 2014; Wang et al., 2016). However, the Training group did not show any greater improvement in math achievement scores compared to the control group. The absence of transfer effect to symbolic math ability after training is consistent with some previous reports (Räsänen et al., 2009; Wilson et al., 2009).

Several aspects of the present study are worth noting. Compared to previous studies, the period of training was much longer and the range of magnitudes used for both training and assessment was much larger. Furthermore, non-numerical visual magnitudes of stimuli were controlled for during both



training and assessment. In addition, different visual interfaces of tasks were used between training vs. assessment to prevent direct practice effects. Our training was conducted in the child's home to improve ecological validity to real-world, educational applications. Based on the analysis of our data, we could not find any evidence in support of training effects that transfer to improvement in any domain of math achievement. The only effect of training observed was improvement in the accuracy of numerosity comparison.

Comparison With Other Training Studies

The results of the present study are in contrast with those reported by some training studies conducted with young children (Wilson et al., 2009; Obersteiner et al., 2013; Hyde et al., 2014; Khanum et al., 2016; Maertens et al., 2016; Park et al., 2016; Sella et al., 2016; Wang et al., 2016). In our study, only

TABLE 4 | The result of mixed 2×2 repeated measures ANOVA on math achievement scores (computerized arithmetic, KNISE-BAAT) with group as between-subject factor and time as within-subject factor.

Dependent variable		Source	SS	df	MS	F	P	η^2
Computerized arithmetic (accuracy)	Within subjects	Time	0.01	1	0.01	0.22	0.64	0.01
		Group \times time	0.01	1	0.01	1.82	0.19	0.04
		Error	0.33	44	0.01			
	Between subjects	Group	0.04	1	0.04	1.18	0.28	0.03
		Error	1.31	44	0.03			
KNISE-BAAT (total score)	Within subjects	Time	1,115.15	1	1,115.15	44.34	0.01	0.50
		Group \times time	0.89	1	0.89	0.04	0.85	0.01
		Error	1,106.60	44	25.15			
	Between subjects	Group	139.16	1	139.16	0.79	0.38	0.01
		Error	7,804.32	44	177.37			

The total score of KNISE-BAAT represents the mean of all subtests.

non-symbolic numerosity comparison performance (but not PAE from symbolic numberline estimation) improved after training without any transfer effects to math achievement. In contrast, Maertens et al. (2016) reported that only the PAE of numberline estimation (but not numerosity comparison) improved significantly more in the training compared to the control group, but both training effects transferred to improvement on pictorially presented (but not symbolic) arithmetic problems in preschoolers. In Hyde et al. (2014), single session practice on both approximate addition and numerosity comparison (but not line length addition or brightness comparison) led to gains in exact, symbolic addition (but not sentence comparison) (Hyde et al., 2014). In Wang et al. (2016), 5-year-old children who were briefly trained to improve their precision of numerosity discrimination showed higher performance on symbolic math (but not vocabulary) compared to the control group. In this study, improvement in children's ANS acuity was brought about by presenting trials in "easy to hard" order, to induce the experience of a sequence of confident problem solving ("confidence hysteresis") which the authors believe leads to enhancement of ability (Odic et al., 2014). In the control groups, trials were presented in the opposite or random orders. However, there were no pre-training assessment of ANS acuity or math ability in Wang et al., (2016), thus it is difficult to rule out pre-training differences in ANS acuity or math ability between groups. Furthermore, some researchers question whether the transfer effects observed in Hyde et al. (2014) or Wang et al. (2016) reflect attentional priming to numerical representations rather than true transfer effects to math improvement, given the brevity of the practices in these two studies (Szűcs and Myers, 2017). Taken together, although it is not possible to definitively state the cause of these discrepancies, possible sources may include differences in how pre- and post-training assessments were made and the duration of training. The present study conducted pre- and post-training assessment of all abilities included in the training program using a separate task designed with a different visual interface. Thus, in the present study, mere practice (or test-retest) effects were minimized in the post-training assessment, making it less likely to see improvement on the outcome ability. It is also possible

that in young children, ANS performance is facilitated by the presence of non-numerical visual magnitudes that are correlated with numerosity (Defever et al., 2013; Szűcs et al., 2013). The present study controlled for the influence of non-numerical visual properties of the stimuli during both training and assessment. Differences in the method by which non-numerical visual magnitudes were controlled across studies may have influenced the discrepancy in the type of cognitive process trained and the resulting assessment of outcome ability. Considering the observation of Szkudlarek and Brannon (2018), it should also be emphasized that individuals (especially young children) with low ability may benefit more from approximate arithmetic training and that transfer effects of training may be specific to certain domains or components of math ability (e.g., informal math skills as opposed to formal math skills).

Factors that May Influence Transfer Effects

Inconsistencies across studies may also be due to differences in the contents of the training across studies. First of all, transfer effects to symbolic math reported from training studies which included symbolic arithmetic practice (e.g., Number Race or Rescue Calcularis) may reflect direct practice effects because the training program itself included symbolic arithmetic practice (Wilson et al., 2006, 2009; Vilette et al., 2010; Kucian et al., 2011; Obersteiner et al., 2013; Sella et al., 2016). Second, training may be less effective when multiple modules are included within a single session. Several training studies which involved practicing a single type of process (e.g., approximate arithmetic, numerosity comparison, or number line estimation) observed significant transfer effects (Park and Brannon, 2013; Hyde et al., 2014; Park and Brannon, 2014; Khanum et al., 2016; Maertens et al., 2016; Park et al., 2016; Au et al., 2018). In contrast, when training involves multiple kinds of training modules (as in our study), transfer effects may be less easily observed, due to increased variability in the effectiveness of each training module across participants. For example, some participants may be relatively more engaged and motivated by module A, while others by module B, and so forth. In such cases, the group average of

the training effect of each module may be reduced by increased individual variability and (by the same reason) the resulting transfer effect may also be washed out, especially if each module is more or less associated with partially different components of the outcome ability. Thus, increased variability in training effects of each module across individuals may have caused the absence of direct improvement on some of the trained tasks (numberline estimation or approximate arithmetic ability, etc.) in the present study, especially given the small sample size.

Based on the observation that training on approximate arithmetic but not numerosity comparison transfers to improvement on symbolic arithmetic in adults, Park and Brannon (2014) hypothesized that cognitive training may have positive transfer effects if training and the outcome ability share *common mental operations* (Park and Brannon, 2014). Alternatively, Hyde et al. (2016) hypothesized that transfer effects may be determined by the *overlap of mental representations* between training and the outcome ability (at least in children). This hypothesis was based on the observation that training effects from both numerosity comparison and approximate addition transferred to improvement in symbolic addition in children (Hyde et al., 2016). The absence of transfer effect to symbolic math ability despite improved ANS acuity in the present study seems to support the idea that transfer effects of training require substantial overlap of mental operations between the trained process and the outcome ability, consistent with Park & Brannon's "Operational Overlap" hypothesis. Taken together, as Hyde et al. (2016) had mentioned as well, we emphasize that finding the answer to the question of which type of basic mathematical training can enhance mathematical cognition requires continued efforts, taking into consideration that factors such as developmental changes and subtle differences in research methodology can critically influence this relationship.

Limitations and Directions for the Future

We acknowledge that it would be better to include another kind of active training (unrelated to basic numerical processing) for the control group. We acknowledge this as a limitation of the present study. If there had been a transfer effect of training on math achievement selectively for the training group, it would have been hard to eliminate the possibility of a placebo (or Hawthorne) effect. However, given the absence of a transfer effect, the lack of a control training program can be thought to be less of a problem in the case of the present study. Furthermore, Szűcs and Myers (2017) emphasizes that it is not meaningful to contrast target-related interventions with target-irrelevant ones (e.g., contrasting math training vs. reading or drawing interventions) or to contrast between two interventions which are not equally engaging, motivating, and intellectually stimulating. Although the present study lacks an active control group, we can at least contrast the efficacy of different types of basic math training based on a within-subjects design, while all training can be considered to be equally engaging and motivating. [All submodules were based on a coherent theme (i.e., animals baking cake, animals selling cake, etc.), user interface (presentation

of colorful cartoons with music), and method of feedback.] Regardless, in future studies, it would be ideal to contrast the effect of cognitive training in the experimental group against a well-matched, alternative form of training for the control group.

The absence of transfer effect in our home-based training study raises the question of whether transfer effects observed within a lab-based setting will generalize to real-world or actual educational applications. Taken together, the results of the present study reveal that (1) only certain kinds of basic numerical ability (in the present case, only ANS acuity) of young children can be improved with training and (2) improvement on ANS acuity does not seem transfer to improvement in math achievement, despite extensive training for 6 weeks, across large ranges of magnitudes.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this manuscript will be made available by the authors, without undue reservation, to any qualified researcher.

AUTHOR CONTRIBUTIONS

NK contributed to designing the pre- and post-training assessments, improving the design of the training program, data collection, and analysis. SJ contributed to designing the training paradigm, data collection, and analysis. SC is the principle investigator who contributed to all aspects of the study from acquiring funding, designing the study, and monitoring all aspects of data collection and analysis. All authors contributed to writing the manuscript.

FUNDING

This research was supported by grants from the National Research Foundation of Korea (2014R1A1A3051034 and 2017R1D1A1B03032115) funded by the Korean government to SC.

ACKNOWLEDGMENTS

We thank all children and their families for their participation. We are also grateful to our lab members and research assistants for their assistance in data collection and analysis. Part of the raw data included in this study is from NK's master's thesis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01775/full#supplementary-material>

REFERENCES

- Agrillo, C., Dadda, M., Serena, G., and Bisazza, A. (2008). Do fish count? Spontaneous discrimination of quantity in female mosquitofish. *Anim. Cogn.* 11, 495–503. doi: 10.1007/s10071-008-0140-9
- Arthur, W. Jr., Tubre, T. C., Paul, D. S., and Sanchez-Ku, M. L. (1999). College-sample psychometric and normative data on a short form of the raven advanced progressive matrices test. *J. Psychoeduc. Assess.* 17, 354–361. doi: 10.1177/073428299901700405
- Au, J., Jaeggi, S. M., and Buschkuhl, M. (2018). Effects of non-symbolic arithmetic training on symbolic arithmetic and the approximate number system. *Acta Psychol.* 185, 1–12. doi: 10.1016/j.actpsy.2018.01.005
- Booth, J. L., and Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Dev. Psychol.* 42:189.
- Codding, R. S., Burns, M. K., and Lukito, G. (2011). Meta-analysis of mathematic basic-fact fluency interventions: a component analysis. *Learn. Disab. Res. Pract.* 26, 36–47. doi: 10.1111/j.1540-5826.2010.00323.x
- Cohen, P. A., Kulik, J. A., and Kulik, C.-L. C. (1982). Educational outcomes of tutoring: a meta-analysis of findings. *Am. Educ. Res. Journal* 19, 237–248. doi: 10.3102/00028312019002237
- De Smedt, B., and Gilmore, C. K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *J. Exp. Child Psychol.* 108, 278–292. doi: 10.1016/j.jecp.2010.09.003
- De Smedt, B., Verschaffel, L., and Ghesquière, P. (2009). The predictive value of numerical magnitude comparison for individual differences in mathematics achievement. *J. Exp. Child Psychol.* 103, 469–479. doi: 10.1016/j.jecp.2009.01.010
- Defever, E., Reynvoet, B., and Gebuis, T. (2013). Task- and age-dependent effects of visual stimulus properties on children's explicit numerosity judgments. *J. Exp. Child Psychol.* 116, 216–233. doi: 10.1016/j.jecp.2013.04.006
- DeWind, N. K., and Brannon, E. M. (2012). Malleability of the approximate number system: effects of feedback and training. *Front. Hum. Neurosci.* 6:68. doi: 10.3389/fnhum.2012.00068
- Dietrich, J. F., Huber, S., and Nuerk, H.-C. (2015). Methodological aspects to be considered when measuring the approximate number system (ANS) – a research review. *Front. Psychol.* 6:295. doi: 10.3389/fpsyg.2015.00295
- Geary, D. C., Hoard, M. K., Nugent, L., and Byrd-Craven, J. (2008). Development of number line representations in children with mathematical learning disability. *Dev. Neuropsychol.* 33, 277–299. doi: 10.1080/87565640801982361
- Gebuis, T., and Reynvoet, B. (2012a). Continuous visual properties explain neural responses to nonsymbolic number. *Psychophysiology* 49, 1649–1659. doi: 10.1111/j.1469-8986.2012.01461.x
- Gebuis, T., and Reynvoet, B. (2012b). The interplay between nonsymbolic number and its continuous visual properties. *J. Exp. Psychol. Gen.* 141, 642. doi: 10.1037/a0026218
- Gersten, R., Chard, D. J., Jayanthi, M., Baker, S. K., Morphy, P., and Flojo, J. (2009). Mathematics instruction for students with learning disabilities: A meta-analysis of instructional components. *Rev. Educ. Res.* 79, 1202–1242. doi: 10.3102/0034654309334431
- Halberda, J., and Feigenson, L. (2008). Developmental change in the acuity of the “Number Sense”: The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Dev. Psychol.* 44:1457. doi: 10.1037/a0012682
- Hyde, D. C., Berteletti, I., and Mou, Y. (2016). Approximate numerical abilities and mathematics: insight from correlational and experimental training studies. *Prog. Brain Res.* 227, 335–351. doi: 10.1016/bs.pbr.2016.04.011
- Hyde, D. C., Khanum, S., and Spelke, E. S. (2014). Brief non-symbolic, approximate number practice enhances subsequent exact symbolic arithmetic in children. *Cognition* 131, 92–107. doi: 10.1016/j.cognition.2013.12.007
- Jang, S., and Cho, S. (2016). The acuity for numerosity (but not continuous magnitude) discrimination correlates with quantitative problem solving but not routinized arithmetic. *Curr. Psychol.* 35, 44–56. doi: 10.1007/s12144-015-9354-6
- Jang, S., and Cho, S. (2018). The mediating role of number-to-magnitude mapping precision in the relationship between approximate number sense and math achievement depends on the domain of mathematics and age. *Learn. Individ. Diff.* 64, 113–124. doi: 10.1016/j.lindif.2018.05.005
- Jordan, N. C., Kaplan, D., Locuniak, M. N., and Ramineni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learn. Disab. Res. Pract.* 22, 36–46. doi: 10.1111/j.1540-5826.2007.00229.x
- Khanum, S., Hanif, R., Spelke, E. S., Berteletti, I., and Hyde, D. C. (2016). Effects of non-symbolic approximate number practice on symbolic numerical abilities in Pakistani children. *PLoS One* 11:e0164436. doi: 10.1371/journal.pone.0164436
- Kroesbergen, E. H., and Van Luit, J. E. (2003). Mathematics interventions for children with special educational needs: a meta-analysis. *Remed. Special Educ.* 24, 97–114. doi: 10.4103/2277-9531.162372
- Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., et al. (2011). Mental number line training in children with developmental dyscalculia. *Neuroimage* 57, 782–795. doi: 10.1016/j.neuroimage.2011.01.070
- Lee, K., and Cho, S. (2017). Magnitude processing and complex calculation is negatively impacted by mathematics anxiety while retrieval-based simple calculation is not. *Int. J. Psychol.* 53, 321–329. doi: 10.1002/ijop.12412
- Leibovich, T., and Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Front. Psychol.* 4:375. doi: 10.3389/fpsyg.2013.00375
- Leibovich, T., Katzin, N., Harel, M., and Henik, A. (2017). From “sense of number” to “sense of magnitude”: the role of continuous magnitudes in numerical cognition. *Behav. Brain Sci.* 40:e164. doi: 10.1017/S0140525X16000960
- Libertus, M. E., Feigenson, L., and Halberda, J. (2011). Preschool acuity of the approximate number system correlates with school math ability. *Dev. Sci.* 14, 1292–1300. doi: 10.1111/j.1467-7687.2011.01080.x
- Locuniak, M. N., and Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *J. Learn. Disabil.* 41, 451–459. doi: 10.1177/0022219408321126
- Lyons, I. M., and Beilock, S. L. (2011). Numerical ordering ability mediates the relation between number-sense and arithmetic competence. *Cognition* 121, 256–261. doi: 10.1016/j.cognition.2011.07.009
- Maertens, B., De Smedt, B., Sasanguie, D., Elen, J., and Reynvoet, B. (2016). Enhancing arithmetic in pre-schoolers with comparison or number line estimation training: does it matter? *Learn. Instr.* 46, 1–11. doi: 10.1016/j.learninstruc.2016.08.004
- Maloney, E. A., Risko, E. F., Preston, F., Ansari, D., and Fugelsang, J. (2010). Challenging the reliability and validity of cognitive measures: The case of the numerical distance effect. *Acta Psychol.* 134, 154–161. doi: 10.1016/j.actpsy.2010.01.006
- Martin, R. B., Cirino, P. T., Sharp, C., and Barnes, M. (2014). Number and counting skills in kindergarten as predictors of grade 1 mathematical skills. *Learn. Individ. Diff.* 34, 12–23. doi: 10.1016/j.lindif.2014.05.006
- Mazzocco, M. M., Feigenson, L., and Halberda, J. (2011). Preschoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS One* 6:e23749. doi: 10.1371/journal.pone.0023749
- Obersteiner, A., Reiss, K., and Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learn. Instr.* 23, 125–135. doi: 10.1016/j.learninstruc.2012.08.004
- Odic, D., Hock, H., and Halberda, J. (2014). Hysteresis affects approximate number discrimination in young children. *J. Exp. Psychol. Gen.* 143:255. doi: 10.1037/a0030825
- Park, J., Bermudez, V., Roberts, R. C., and Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in preschoolers. *J. Exp. Child Psychol.* 152, 278–293. doi: 10.1016/j.jecp.2016.07.011
- Park, J., and Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychol. Sci.* 24, 2013–2019. doi: 10.1177/0956797613482944
- Park, J., and Brannon, E. M. (2014). Improving arithmetic performance with number sense training: an investigation of underlying mechanism. *Cognition* 133, 188–200. doi: 10.1016/j.cognition.2014.06.011
- Park, K., Kim, K., Song, Y., Jeong, D., and Jeong, I. (2008). *KISE-Basic Academic Achievement Tests (KISE-BAAT)*. Ansan: Korea National Institute for Special Education.
- Park, Y., and Cho, S. (2016). Developmental changes in the relationship between magnitude acuities and mathematical achievement in elementary school children. *Educ. Psychol.* 37, 873–887. doi: 10.1080/01443410.2015.1127332

- Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., et al. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition* 116, 33–41. doi: 10.1016/j.cognition.2010.03.012
- Pica, P., Lemer, C., Izard, V., and Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science* 306, 499–503. doi: 10.1126/science.1102085
- Räsänen, P., Salminen, J., Wilson, A. J., Aunio, P., and Dehaene, S. (2009). Computer-assisted intervention for children with low numeracy skills. *Cogn. Dev.* 24, 450–472. doi: 10.1016/j.cogdev.2009.09.003
- Rousselle, L., and Noël, M.-P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs non-symbolic number magnitude processing. *Cognition* 102, 361–395. doi: 10.1016/j.cognition.2006.01.005
- Sasanguie, D., De Smedt, B., Defever, E., and Reynvoet, B. (2012). Association between basic numerical abilities and mathematics achievement. *Br. J. Dev. Psychol.* 30, 344–357. doi: 10.1111/j.2044-835X.2011.02048.x
- Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., and Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number-space mappings: What underlies mathematics achievement? *J. Exp. Child Psychol.* 114, 418–431. doi: 10.1016/j.jecp.2012.10.012
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2016). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372
- Sella, F., Tressoldi, P., Lucangeli, D., and Zorzi, M. (2016). Training numerical skills with the adaptive videogame “The Number Race”: A randomized controlled trial on preschoolers. *Trends Neurosci. Educ.* 5, 20–29. doi: 10.1016/j.tine.2016.02.002
- Siegler, R. S., and Booth, J. L. (2004). Development of numerical estimation in young children. *Child Dev.* 75, 428–444. doi: 10.1111/j.1467-8624.2004.00684.x
- Siegler, R. S., and Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income preschoolers’ numerical understanding. *J. Educ. Psychol.* 101:545. doi: 10.1037/a0014239
- Starr, A., Libertus, M. E., and Brannon, E. M. (2013). Number sense in infancy predicts mathematical abilities in childhood. *Proc. Natl. Acad. Sci. U.S.A.* 110, 18116–18120. doi: 10.1073/pnas.1302751110
- Szkudlarek, E., and Brannon, E. M. (2018). Approximate Arithmetic Training Improves Informal Math Performance in Low Achieving Preschoolers. *Front. Psychol.* 9:606. doi: 10.3389/fpsyg.2018.00606
- Szűcs, D., and Myers, T. (2017). A critical analysis of design, facts, bias and inference in the approximate number system training literature: A systematic review. *Trends Neurosci. Educ.* 6, 187–203. doi: 10.1016/j.tine.2016.11.002
- Szucs, D., Nobes, A., Devine, A., Gabriel, F. C., and Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Front. Psychol.* 4:444. doi: 10.3389/fpsyg.2013.00444
- Vilette, B., Mawart, C., and Rusinek, S. (2010). L’outil «estimateur», la ligne numérique mentale et les habiletés arithmétiques. *Pratiques Psychol.* 16, 203–214. doi: 10.1016/j.prps.2009.10.002
- Wang, J., Odic, D., Halberda, J., and Feigenson, L. (2016). Changing the precision of preschoolers’ approximate number system representations changes their symbolic math performance. *J. Exp. Child Psychol.* 147, 82–99. doi: 10.1016/j.jecp.2016.03.002
- Wilson, A. J., Dehaene, S., Dubois, O., and Fayol, M. (2009). Effects of an adaptive game intervention on accessing number sense in low-socioeconomic-status kindergarten children. *Mind Brain Educ.* 3, 224–234. doi: 10.1111/j.1751-228X.2009.01075.x
- Wilson, A. J., Revkin, S. K., Cohen, D., Cohen, L., and Dehaene, S. (2006). An open trial assessment of “The Number Race”, an adaptive computer game for remediation of dyscalculia. *Behav. Brain Funct.* 2:20.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Kim, Jang and Cho. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Symbolic Number Comparison Is Not Processed by the Analog Number System: Different Symbolic and Non-symbolic Numerical Distance and Size Effects

Attila Krajcsi^{1*}, Gábor Lengyel² and Petia Kojouharova^{3,4}

¹ Cognitive Psychology Department, Institute of Psychology, Eötvös Loránd University, Budapest, Hungary, ² Department of Cognitive Science, Central European University, Budapest, Hungary, ³ Doctoral School of Psychology, Eötvös Loránd University, Budapest, Hungary, ⁴ Research Centre for Natural Sciences, Institute of Cognitive Neuroscience and Psychology, Hungarian Academy of Sciences, Budapest, Hungary

OPEN ACCESS

Edited by:

Jingguang Li,
Dali University, China

Reviewed by:

Wei Liu,
Yunnan Nationalities University, China
Thomas J. Faulkenberry,
Tarleton State University,
United States

*Correspondence:

Attila Krajcsi
krajcsi@gmail.com

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 20 November 2017

Accepted: 25 January 2018

Published: 09 February 2018

Citation:

Krajcsi A, Lengyel G and
Kojouharova P (2018) Symbolic
Number Comparison Is Not
Processed by the Analog Number
System: Different Symbolic and
Non-symbolic Numerical Distance and
Size Effects. *Front. Psychol.* 9:124.
doi: 10.3389/fpsyg.2018.00124

HIGHLIGHTS

- We test whether symbolic number comparison is handled by an analog noisy system.
- Analog system model has systematic biases in describing symbolic number comparison.
- This suggests that symbolic and non-symbolic numbers are processed by different systems.

Dominant numerical cognition models suppose that both symbolic and non-symbolic numbers are processed by the Analog Number System (ANS) working according to Weber's law. It was proposed that in a number comparison task the numerical distance and size effects reflect a ratio-based performance which is the sign of the ANS activation. However, increasing number of findings and alternative models propose that symbolic and non-symbolic numbers might be processed by different representations. Importantly, alternative explanations may offer similar predictions to the ANS prediction, therefore, former evidence usually utilizing only the goodness of fit of the ANS prediction is not sufficient to support the ANS account. To test the ANS model more rigorously, a more extensive test is offered here. Several properties of the ANS predictions for the error rates, reaction times, and diffusion model drift rates were systematically analyzed in both non-symbolic dot comparison and symbolic Indo-Arabic comparison tasks. It was consistently found that while the ANS model's prediction is relatively good for the non-symbolic dot comparison, its prediction is poorer and systematically biased for the symbolic Indo-Arabic comparison. We conclude that only non-symbolic comparison is supported by the ANS, and symbolic number comparisons are processed by other representation.

Keywords: Analog Number System, number comparison, Weber's law, diffusion model, symbolic numbers

REPRESENTATION BEHIND SYMBOLIC NUMBER PROCESSING

Analog Number System

In their seminal work Moyer and Landauer (1967) described that in an Indo-Arabic single digit number comparison task the performance is worse (i.e., reaction time is slower and error rate is higher) when the difference between the two numbers is relatively small (numerical distance effect) or when the numbers are relatively large (numerical size effect). They proposed that the effects are the expression of a general ratio-based effect in which number pairs with smaller ratio are harder to process. This ratio-based performance was thought to be the result of a simple representation working according to Weber's law, termed the Analog Number System (ANS), similar to the representations working behind simple physical feature comparison tasks. Since then, the ratio-based performance (usually measured only with the distance effect) is thought to be the signal of a noisy analog representation working in the background.

The ratio-based performance was also specified with quantitative descriptions. Originally, Moyer and Landauer (1967) demonstrated that the reaction time pattern can be described appropriately with a function used at that time in physical property comparison tasks: a $K \times \log(\text{large_number}/(\text{large_number}-\text{small_number}))$ function correlates well with the measured reaction time, $r = 0.75$. Later, more precise mathematical descriptions were offered (see Dehaene, 2007 for an extensive mathematical description of the model). According to one of the implementations of these descriptions, the numbers are stored as noisy representation following a Gaussian distribution, and the noise is proportional to the value of the number. This increasing noise can produce the ratio-based performance. For example, the overlap between the representations of two numbers predicts the error rate in a comparison task, or more generally, this overlap predicts the difficulty of the task, expressed as drift rate in the diffusion model (see more details in the Methods section). (This proportionally increasing noise can also be implemented in a logarithmic representation with constant noise on a logarithmic scale).

The ANS is supposed to work behind any number comparison, independent of the notation of the numbers (Dehaene, 1992; Nieder, 2005; Piazza, 2010), because the same ratio-based performance can be observed behind symbolic and non-symbolic tasks (Moyer and Landauer, 1967; Dehaene, 2007), and because overlapping brain areas are activated in symbolic and non-symbolic number processing (Eger et al., 2003; Nieder, 2005). Although there could be differences between the symbolic and non-symbolic number processing, and even there could be two different representations working with different sensitivity (i.e., Weber fraction), both of these stimuli are processed by the same *type* of representations, which representations work according to Weber's law, producing a ratio-based performance (Dehaene, 2007; Piazza, 2010).

The common mechanism and the strong relation between symbolic and non-symbolic processing is also reflected by several findings showing that, for example, the sensitivity of the ANS measured in a non-symbolic dot comparison task correlates with

symbolic math achievement (Halberda et al., 2008), or training non-symbolic number processing improves the symbolic number processing (Park and Brannon, 2013). To summarize, it is widely supposed that number processing is supported by a noisy, analog representation, working according to Weber's law, and therefore producing a ratio-based performance in comparison tasks. Also, this type of mechanism works behind both symbolic and non-symbolic number processing, as reflected by many similarities and relations between symbolic and non-symbolic numerical tasks.

Different Symbolic and Non-symbolic Number Processing

However, there are increasing number of findings in the literature suggesting that the symbolic and non-symbolic number processing is not backed by the same representation or by the same type of representations. For example, it has been shown that performance of the symbolic and non-symbolic number comparison tasks do not correlate in children (Holloway and Ansari, 2009; Sasanguie et al., 2014). As another example, while former studies found that common brain areas are activated by both symbolic and non-symbolic stimuli (Eger et al., 2003; Piazza et al., 2004), later works with more sensitive methods found only notation-specific activations (Damarla and Just, 2013; Bulthé et al., 2014, 2015). In another fMRI study, the size of the symbolic and non-symbolic number activations did not correlate, and more importantly, the activation for the symbolic number processing seemed to be discrete and not analog (Lyons et al., 2015a). According to an extensive meta-analysis, while it was repeatedly found that the simple number comparison task (the supposed index for the sensitivity of the ANS) correlates with mathematical achievement, it seems that non-symbolic comparison correlates much less with math achievement, than symbolic comparison (Schneider et al., 2017). In another example, Noël and Rousselle (2011) found that while older than 9- or 10-year-old children with developmental dyscalculia (DD) perform worse in both symbolic and non-symbolic tasks than the typically developing children, younger children with DD perform worse than control children only in the symbolic tasks, but not in the non-symbolic tasks, meaning that the deficit is more strongly related to the symbolic number processing, and the impaired non-symbolic performance is only the consequence of the symbolic processing problems. See a more extensive review of similar findings in Leibovich and Ansari (2016). All of these findings are in line with the present proposal, suggesting that symbolic and non-symbolic numbers are processed by different systems.

Additionally, there are a few alternative models that are in line with these later findings showing that symbolic and non-symbolic number processing is not backed by the same representation or by the same type of systems. In a connectionist model of symbolic number processing, the model successfully explains many phenomena the ANS model cannot handle (Verguts et al., 2005; Verguts and Van Opstal, 2014). Although this model is interpreted as a version of the ANS (Verguts

and Fias, 2004; Dehaene, 2007), critically, it does not show the defining feature of the ANS: the model does not produce inherently the ratio-based performance, instead, introduction of the uneven frequency of the digits is necessary to produce the size effect (Verguts and Fias, 2004; Verguts et al., 2005). Thus, the model proposes different type of mechanisms for symbolic and non-symbolic number processing. Another model assumes that primitives (simple representational units) are stored in the long term memory only for the digits (numbers between 0 and 9) (Pinhas and Tzelgov, 2012), but not for other values (Kallai and Tzelgov, 2009; Tzelgov et al., 2009), suggesting a symbolic-only representation. In a third model it was proposed that symbolic numbers can be stored in a Discrete Semantic System (DSS), similar to the mental lexicon or a semantic network. In this system numbers are represented by nodes, and the connections of the nodes reflect the semantic relations of the nodes mostly directed by the numerical distance of the number pairs (Krajcsi et al., 2016). The distance effect might be originated in the semantic relation of the nodes, as was seen in the similar semantic distance effect in a picture naming task (Vigliocco et al., 2002). The numerical size effect could be rooted in the fact that smaller numbers are more frequent than larger numbers (Dehaene and Mehler, 1992), and more frequent numbers can be processed more easily. The DSS model can be easily extended to account for symbolic numerical interference effects as well (Proctor and Cho, 2006; Leth-Steensen et al., 2011; Patro et al., 2014). Thus, the DSS can account for symbolic numerical effects, independent of the non-symbolic number processing.

Importantly, in the DSS account a performance pattern similar to the ANS model can be offered. For example, it is possible that the reaction time could be proportional to the sum of the linear distance effect and the size effect originated in the frequency of the values, which in turn is related to the power of those values (see the justification for this function and similar possibilities in Krajcsi et al., 2016). **Figure 1** shows two possible implementations of the ANS and the DSS models, and it reveals that the DSS model might generate a very similar pattern to the one supposed by the ANS model (the correlation of the two presented performance predictions is -0.89).

The similarity of the ANS and the DSS model predictions means that the DSS model could be potentially an appropriate alternative explanation for the observed distance and size effects. Even more importantly, this means that former works investigating whether the ANS model is correct might have found high correlation between the ANS model and the observed performance either because the ANS model is correct, or because it is the DSS model that is correct, and as the ANS model prediction correlates highly with the DSS model prediction, the correlation between the ANS prediction and the performance was only illusory.

To summarize, an increasing body of evidence indicates that symbolic and non-symbolic numbers might be processed by different types of representations, and there could be appropriate alternative models to explain symbolic number processing, which may also question the suitability of former tests.

The Aim of the Study

The aim of the present study is to test the appropriateness of the ANS model in comparison tasks more extensively. The appropriateness of the ANS model for both symbolic and non-symbolic notations have been investigated several times, finding that the prediction of the ANS model is similar to the observed performance (Moyer and Landauer, 1967; e.g., Dehaene, 2007). Former studies usually investigated the goodness of fit of the ANS model for the observed performance. However, these former tests are insufficient, because similarity between the ANS model prediction and the observed performance may be caused by alternative models with similar predictions, such as the DSS model. For example, it is possible that in the Moyer and Landauer (1967) study, the $r = 0.75$ correlation between the observed reaction time and the ANS model prediction is the result of a stronger than $r = 0.75$ correlation between the DSS model prediction and the observed performance, and the strong correlation between the DSS model and the ANS model predictions (e.g., $r = -0.89$). Therefore, it is not enough to show that the ANS model's prediction is similar to the observed data, but a more extensive test is needed.

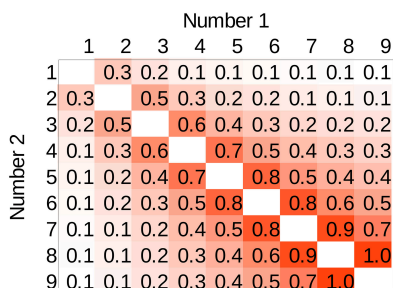
Here we test the appropriateness of the ANS model by investigating whether the ANS model can explain both symbolic and non-symbolic comparison tasks equally well, or whether there are critical differences between symbolic and non-symbolic comparison tasks.¹ If the ANS account is correct, then one should expect that the ANS model can describe both symbolic and non-symbolic equally well, as suggested repeatedly in the literature (Moyer and Landauer, 1967; Eger et al., 2003; Nieder, 2005; Dehaene, 2007). However, if there are differences between the symbolic and non-symbolic notations, one might suppose that the ANS can describe the non-symbolic comparison appropriately, in line with the fact that non-symbolic stimuli are visual-perceptual as other physical properties processed by other representations working according to Weber's law (Moyer and Landauer, 1967; Dakin et al., 2011; Gebuis and Reynvoet, 2012; Stoianov and Zorzi, 2012), while the ANS model cannot account for the symbolic comparison, as suggested by the alternative symbolic number processing models.

One might question whether this type of test is meaningful, because symbolic and non-symbolic comparison do not necessarily work in the same way, even if the ANS model is correct. For example, there could be additional notation-specific mechanisms that could change behavioral performance, therefore, one cannot expect that the two notations should show the same performance pattern. However, if someone believes that there could be additional components that might influence the behavioral performance, then one must also question whether the findings suggesting ratio-based performance in any comparisons

¹ A similar investigation of symbolic and non-symbolic comparisons testing against the ANS model was done by Dehaene (2007), however, in that study multi-digit Indo-Arabic numbers were utilized. When comparing multi-digit symbolic numbers one might process the numbers power by power, and holistic ANS processing of the number cannot be guaranteed (Hinrichs et al., 1982; Poltrock and Schwartz, 1984; Krajcsi and Szabó, 2012; Huber et al., 2016), therefore, in such a test multi-digit symbolic numbers should be avoided. The present work utilizes only single digit Indo-Arabic numbers.

Analogue Number System (ANS) model

$$RT = a \times \log(\text{large}/\text{distance}) + b$$

**Discrete Semantic System (DSS) model**

$$RT = a_1 \times \text{distance} + a_2 \times (x_1^{-1} + x_2^{-1}) + b$$

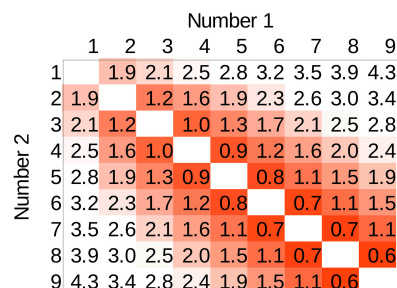


FIGURE 1 | Two possible predictions for the reaction time pattern by the ANS model (based on Crossman, 1955; Moyer and Landauer, 1967) and the DSS model (based on Krajcsi et al., 2016) in a comparison task. The prediction of the models on a full stimulus space in a number comparison task of numbers between 1 and 9. Number 1 and 2 are the two values to be compared. White denotes fast responses, red denotes slow responses (note that numerically the ANS function increases, and the DSS function decreases toward the high ratio, but the direction is irrelevant in the linear fit below). The distance effect can be seen as the gradual change when getting farther from the top-left bottom-right diagonal, and the size effect is seen as the gradual change from top-left to bottom-right. Notations: large: larger number; distance: distance between the two numbers; x_1 and x_2 : the two numbers; a , a_1 , a_2 and b are free parameters. In the figures the parameters a and a_2 are set to 1, a_1 is 0.4, and parameter b is set to 0. See also the Methods section for the interpretation of these heat map graphs.

are valid: even if ratio-based performance is observed, the contribution of the hypothesized additional components should be removed, and if that additional component is unspecified, then nothing could be known about the real mechanism in the background. According to this view, the findings of Moyer and Landauer (1967) or any similar results cannot lead to the conclusion that a ratio-based mechanism is working in the background. Overall, one can believe that the current test is invalid, but at the same time it should also be supposed that all tests demonstrating a ratio-based comparison performance are invalid. Even if this viewpoint might seem unusual, it still could be valid. In this case, another types of tests should be found (see for alternative approaches for these tests in Krajcsi et al., 2016; Krajcsi, 2017). But if one thinks that the works that have proposed that ratio-based performance were valid, the present test should be considered to be valid, too.

In the present work we systematically examine whether ANS predicts both symbolic and non-symbolic number comparison performance equally well. Specifically, we examine (1) whether the error rates can be described equally well by the functions derived from the ANS model, (2) whether the reaction time pattern of the two notations fit each other linearly, and (3) whether the diffusion model drift rates of the two notations can be described by the same analog representation. According to the widely accepted version of the ANS model, the model should predict any comparison equally well, because the same ANS-type mechanism processes any numbers independent of their notations. On the other hand, the alternative views might suggest that the ANS should work relatively well only for the non-symbolic notation, but it should work relatively poorly for symbolic notation, because symbolic precise numbers are processed by other mechanisms. Finally, from a methodological point of view, it is also possible that the difference between the ANS and the alternative models is much smaller than the typical noise in the measured data, thus, even if there are differences

between the symbolic and non-symbolic comparisons, the signal-to-noise ratio is not high enough to reveal the difference. For this reason only different behavioral patterns of symbolic and non-symbolic comparisons can be conclusive, supporting the alternative accounts, while lack of difference between the symbolic and non-symbolic comparisons could be either due to the correct ANS description or due to the lack of statistical power.

METHODS

Participants compared Indo-Arabic numbers in one condition, and they compared dot arrays in another condition. In both conditions error rate and reaction time were measured.

Stimuli and Procedure

In a trial two numbers were visible on the left and on the right sides of the screen, and participants had to choose the larger one by pressing one of the two response keys. The stimuli were visible until key press. The response was followed by an empty screen for 500 ms, then the next trial started.

In the Indo-Arabic condition the numbers were between 1 and 9, to avoid multi-digit numbers (see footnote 1 for more details). All possible pairings of those values were presented, except ties, resulting in 72 possible pairs. All pairs were presented 10 times, resulting 720 trials in the condition. The order of the trials was randomized.

In the dots condition it is not appropriate to use the same 1–9 range as in the Indo-Arabic condition, because sets with less than five objects can be enumerated fast, which fast enumeration is termed subitizing (Kaufman et al., 1949). Subitizing is not an ANS directed process (Revkin et al., 2008), but it is most probably based on pattern detection (Mandler and Shebo, 1982; Krajcsi et al., 2013). Therefore, to measure the ANS based dot estimation, the 1–4 range should be avoided. One option could be to use only the numbers between 5 and 9, however, this

solution would considerably decrease the stimulus space. Instead, another solution was applied: it was not the 1–9 range itself that was kept in the dot condition, but the ratios of the 1–9 range. Because according to the ANS model, it is the ratio of the numbers that determines performance, changing the values should not change the performance if the ratios of the values are kept. Therefore, to avoid the 1–4 range, and to keep the critical ratio-based feature at the same time, all numbers between 1 and 9 were multiplied by 5, resulting in a number range between 5 and 45.² In an array of dots, black and white dots in random positions were shown against a gray background (Dakin et al., 2011), thus, the luminance of the stimuli was not informative about the numerosity. Dots of an array were drawn randomly in a $2 \times 2^\circ$ area, with a dot diameter of 0.2° , therefore, density and convex hull correlated with the numerosity. Although our stimuli do not control all perceptual features that might influence the perceived numerosity, in the current test, non-numerical influence of the decision process is less relevant, because the ANS model suggests that number comparison is handled by an analog system that could be used in any continuous physical feature comparison (Moyer and Landauer, 1967; Dehaene, 2007), hence, in a general sense, any continuous physical feature comparison working according to the Weber's law could be an appropriate task in our test. Additionally, a mixture of visual ratio-based performance and numerosity ratio-based performance should also produce an approximately ratio-based performance, as reflected in the similar psychometric functions of visual comparison and numerical comparison tasks. Therefore, the simple and limited visual control of the stimuli is appropriate for the aim of the current test.³ As in the Indo-Arabic condition, all possible pairs were presented 10 times, resulting in 720 trials in the condition. The order of the trials was randomized.

The order of the conditions was counterbalanced across participants.

Participants

Twenty-four university students gave informed consent and participated in the study for partial credit course.⁴ Four participants were excluded, because their error rates were higher than 1.5 standard deviation + mean error rates at least in one of the conditions (6% in the Indo-Arabic condition and 15% in the

dots condition). Among the remaining 20 participants there were 4 males, the age range was 19–24 years, with a mean of 21.0 years.

Analysis Methods

Figures Used in the Results Section

To explore the results in more detail, instead of showing the distance and size effects in the traditional way, the full stimulus space is displayed. The left of **Figure 1** shows how an ANS predicted pattern would look like. Rows and columns denote the two numbers to be compared, and the cells include the performance for a specific number pair. In this figure larger values (on an arbitrary scale) and darker colors denote worse performance.

To relate the current figures to the more widely known effects, in **Figure 2** some “pure” components of the typical patterns can be seen. Distance effect is displayed as the distance from the top-left and bottom-right diagonal, and size effect is displayed as the distance from the top-left corner along a top-left and bottom-right diagonal. Both effects can also be seen in **Figure 1**, because the task is harder close to the top-left and bottom-right diagonal (distance effect) and because the task is harder toward the bottom-right corner (size effect). Traditionally, distance and size effects are computed as calculating the mean performance of the cells with the same distance or size values. Sometimes the end effect is also observable (**Figure 2**, when performance is better with the largest or smallest numbers of the range used in the task (Scholz and Potts, 1974; Balakrishnan and Ashby, 1991; Sathian et al., 1999; Piazza et al., 2002).

These more detailed figures are more appropriate to explore the performance, because (1) any effects that are slightly deviating from the traditional distance and size effects are more visible, and (2) due to the large number of cells systematic patterns can be identified as reliable effects instead of being a random noise, thus, a continuous change in the pattern might signal a specific effect even without statistical hypothesis tests, and random irregularities can be identified as noise.

Error Rate, Reaction Time, and Drift Rate Analysis

Error rate

In psychophysics, specific functions can be found that describe the error rates in a comparison task based on the stimulus intensities and the Weber ratio (Kingdom and Prins, 2010). These functions are also used in the numerical literature (Dehaene, 2007), serving as a firm base to characterize the ANS model prediction. The functions stem from the model summarized in the Introduction, suggesting that error rate is proportional to the overlap of Gaussian noisy representations. In our analysis we used the function described in Dehaene (2007 Equation 10), which supposes a linear scaling in the ANS,

$$p_{\text{correct}}(n_1, n_2) = \int_0^{+\infty} e^{-\frac{1}{2} \left(\frac{x-(r-1)}{w\sqrt{1+r^2}} \right)^2} \frac{dx}{\sqrt{2\pi w\sqrt{1+r^2}}}$$

where n_1 and n_2 are the two numbers to be compared, r is the ratio of the larger and the smaller number, and w is the

²One might raise that this way the two notations do not use the same number ranges, consequently, the two conditions are not comparable. However, it is important to highlight that the current work tests the ANS model, and in this specific test any modifications that are in line with the ANS model are appropriate. If the ratio-based transformations were not allowed, it would already mean that the ANS model is incorrect, therefore, no further test would be needed.

³Similar to the reasoning in the previous footnote, we take advantage of the fact that this work is an ANS test, and any addition that is in line with the ANS model, is allowed. If one questions that the number-based comparison performance has different properties than physical feature comparison performance, then the ANS model itself is questioned, therefore, no further test would be needed.

⁴Because it is impossible to tell what effects sizes can be expected, or even what properties could differ between the two notations, it is not possible to specify an appropriate sample size in advance. Approximately 25 participants were set as a convenient sample size where the most important effects are firmly observable, but no reasonable prediction could be made regarding the reliability of yet unknown differences between the notations.

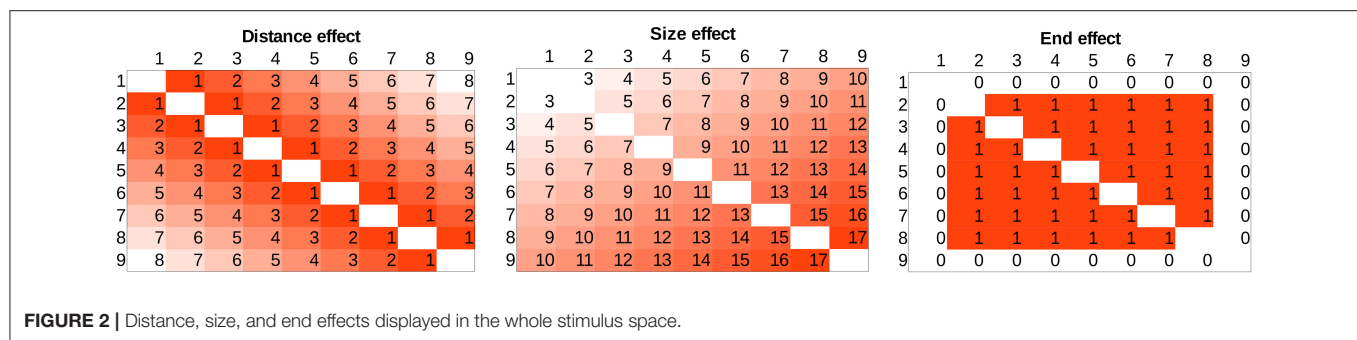


FIGURE 2 | Distance, size, and end effects displayed in the whole stimulus space.

Weber ratio. According to the model this function should work with both symbolic and non-symbolic comparison, although the Weber fraction could be different (Dehaene, 2007). In our analysis the error rates predicted by the specified function above were fit to the group mean of the error rates for both symbolic and non-symbolic comparison for the whole stimulus space.

Reaction time

Current models are not straightforward about the reaction time prediction, and former descriptions (such as used in Moyer and Landauer, 1967) are incorrect from the viewpoint of the current models. Still, to test whether former pieces of evidence were used correctly to support the ANS model, we analyzed the reaction time data.

In the last decades the diffusion model (see the Drift rate section in the Analysis methods for details) became a successful and an increasingly popular tool to describe the reaction time of simple decision processes, including psychophysics comparison tasks. However, earlier works used some simpler models to describe the comparison tasks (Crossman, 1955; Welford, 1960; Moyer and Landauer, 1967). From the perspective of the diffusion models these early descriptions are incorrect, because, for example, they did not consider the Weber ratio of the processing system. Still, because evidence using these methods was considered to support the ANS model, in this detailed exploration we also investigate whether these historical tools can support the idea that the ANS processes both symbolic and non-symbolic numbers.

In these early models, there was no clear consensus about the exact function that could describe the reaction time pattern. Psychophysics was more interested in error rates close to the threshold, and much less work investigated the reaction time far from the threshold (Crossman, 1955). For example, the seminal work by Moyer and Landauer (1967) used the $K \times \log(\text{large_number}/\text{distance})$ function⁵, referring to the Welford (1960) paper, which in turn relied on Crossman (1955), however no straightforward solution was proposed then.

Although it is not easy to specify the function that was thought to describe correctly the reaction time pattern of comparison

tasks, we can avoid this problem. First, as all models agree that dot comparison is handled by the ANS, dot comparison can be considered as the empirical specification of the required function. Second, in the early models, the specific functions could be fitted linearly to the reaction time: the model can be multiplied by a parameter to fit to the time scale of the comparison process, and a parameter can be added to account for the non-decision time. Moyer and Landauer (1967) also used this method implicitly: they reported Pearson product-moment correlation coefficient between the model and the data, which relies on simple linear regression. The linear transformation between the functions and the data means that the measured patterns should be linear transformations of other measured patterns, too. To summarize, according to the analysis methods of early works, the reaction time patterns of different notations are linear transformations of each other. To test this supposition, we fit the dot comparison reaction time pattern to the Indo-Arabic reaction time pattern. Because both dot and Indo-Arabic comparison data include noise, R^2 is not a suitable index to evaluate the similarity of the patterns. However, looking at the residuals can be more informative: if the two patterns readily fit, then only random noise is expected in the residuals. If, on the other hand, the two patterns differ in shape, then the residuals should show a systematic pattern.

It could be possible to have a more appropriate reaction time pattern with applying the diffusion models (see the next part for details), however, to our knowledge there is no clear consensus among others about the functional relationship between the drift rate and the representational overlap, consequently, the reaction time performance could not be specified easily.

Because the reaction time analysis applied here follows the reasoning of the early analysis, the current results cannot be considered as a reliable test of the ANS model, but we examine whether evidence offered formerly really support the common mechanism for symbolic and non-symbolic number processing.

Drift rate

In the recent decades, the diffusion model and related models became increasingly popular to describe simple decision processes (Smith and Ratcliff, 2004; Ratcliff and McKoon, 2008). These models can recover background parameters directing both error rates and reaction times more sensitively. In the diffusion model, decision is based on a gradual accumulation of evidence offered by perceptual and other systems. Decision is made when

⁵In the Moyer and Landauer (1967) paper the $K \log(\text{large_number}/\text{large_number} - \text{small_number})$ function can be found, without the necessary brackets around the large-small term, but most probably the calculation was performed with the correct function.

appropriate amount of evidence is accumulated. Reaction time and error rates partly depend on the quality of the information (termed the drift rate) upon which the evidence is built. Larger drift rate usually results in faster and less erroneous responses. Drift rates are more informative than the error rate or reaction time in themselves, because drift rates reveal the sensitivity of the background mechanisms more directly (Wagenmakers et al., 2007). Importantly for our analysis, observed reaction time and error rate parameters can be used to recover the drift rates (Ratcliff and Tuerlinckx, 2002; Wagenmakers et al., 2007). The drift rates recovered from the behavioral data then can be used to investigate whether they are in line with the prediction of the ANS model.

In the ANS model, like in the case of the error rates, difficulty of the comparison of two properties might depend on the overlap of the two Gaussian random variables: larger overlap leads to worse performance (see the detailed mathematical description in Dehaene, 2007). In the diffusion model framework it is supposed that in a comparison task the drift rate depends purely on the overlap of the two random variables (Palmer et al., 2005; Dehaene, 2007)⁶.

To recover the drift rates for all number pairs in the two notations, the EZ diffusion model was applied, which can be used when the number of trials per cells is relatively small (Wagenmakers et al., 2007). Although this method has several limitations compared to more complex methods (Ratcliff and Tuerlinckx, 2002), (a) all other methods have different limitations, (b) according to current models, the constraints applied in the EZ-diffusion model might not influence the recovered drift rates essentially (although many aspects of the diffusion models are not known yet), and (c) in another numerical task analysis it was found that other tested diffusion models reveal the same pattern as the EZ diffusion model analysis (Kamienkowski et al., 2011). For edge correction we used the half trial solution (see the exact details about edge correction in Wagenmakers et al., 2007). The scaling within-trials variability of drift rate was set to 0.1 in line with the tradition of the diffusion analysis literature.

In the analysis we investigated (a) whether the recovered drift rates are proportional to task difficulty and whether drift rates tend to 0 as the task difficulty increases, and (b) whether drift rates depend purely on the supposed representational overlap, as supposed by the ANS model. As in the case of the error rates,

according to the ANS model, these properties should be present in both symbolic and non-symbolic comparisons (Dehaene, 2007).

RESULTS AND DISCUSSION

Mean Error Rates and Mean Reaction Times

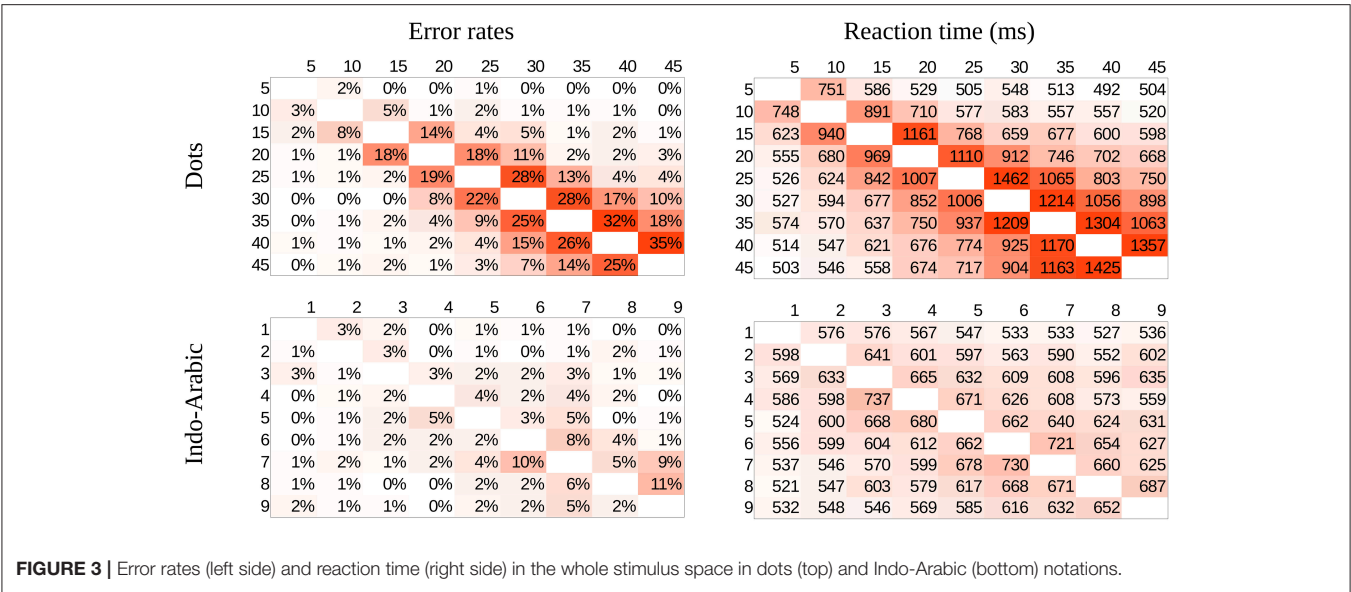
Mean error rates and mean reaction times for correct responses were calculated for all number pairs for all participants in the two notations, then mean values across participants were computed (Figure 3). In both notations distance and size effects are visible, the patterns of the two notations seem similar, and based on first visual inspection the patterns could be in line with both the ANS model and the DSS model predictions.

Two Weber Ratios

The error rate results also revealed that the dot comparison is more erroneous than the Indo-Arabic comparison (the mean of the cells are 6.7% for dot notation and 2.0% for Indo-Arabic notation). On one hand, this result is hardly surprising: even common sense would suggest that the exact symbolic comparison is more precise than an imprecise dot array estimation. On the other hand it raises some nontrivial questions. If both types of comparisons are supported by the same representation, how is it possible that the two types of comparisons show radically different error rates and reaction times?

Because the ANS model suggests that the underlying representation works according to Weber's law, a reasonable idea is that the two notations are supported by different Weber ratios: for the Indo-Arabic comparison a more precise, low value is used, while for the dot array comparison a more imprecise, high value is applied. Dehaene (2007) also suggests that the different Weber ratios can be implemented in different neural cells, similar to the simulation in a connectionist model (Verguts and Fias, 2004). In this connectionist model an ANS-like layer represents the values, which layer works according to Weber's law, and after introducing symbolic notation to the network, the nodes of the number layer become more precise. While this explanation about the two Weber ratios seems compelling, there are some problems that are not trivial to solve. (1) Even if the Weber ratio is relatively small, soon it will reach a ratio in which the noise and the error rates will be too high to complete precise comparisons successfully. However, humans can compare numbers with any precision, which would require an unreasonably small Weber ratio. If one argues that there should be a supplementary mechanism that could help with the very small ratio number pairs, then why is its contribution practically invisible as suggested by the ANS model implicitly (i.e., if the Indo-Arabic comparison performance can be predicted precisely by the ANS model, then no other mechanism should have a major contribution to the measured performance)? (2) Actually, as already discussed in the Introduction, the Verguts model cannot be considered as an ANS model, because after introducing the symbolic numbers, the number layer cannot produce the size effect, violating the ratio-based performance which is a defining feature of the ANS

⁶According to the current models, it is only the drift rate that is relevant in comparison performance (Palmer et al., 2005; Dehaene, 2007). For example, non-decision time is not relevant in the distance effect, because it is not related to the comparison phase (Dehaene, 1996). Similarly, decision threshold is believed to be mainly modulated by the speed-precision instruction (Smith and Ratcliff, 2004) and not by the properties of the stimuli of specific trials. Although it is rare that other than the mean of the performance is investigated in a study, Rouder et al. (2005) measured reaction time properties as parameters of a 3-parameter Weibull distribution. They found that distance effect modified the scale parameter, but not the shape or location parameters. While the relation of the diffusion model generated performance distribution and the Weibull distribution is not fully understood (Rouder et al., 2005), drift rate change of the diffusion model can result in scale parameter change but not in shape or location parameter changes in the 3-parameter Weibull distribution (Rouder et al., 2005), in line with the idea that it is the drift rate that is related to the numerical distance effect.



model (Verguts and Fias, 2004), and only the addition of number frequency could restore the size effect in the model (Verguts et al., 2005), thus, the model cannot work according to Weber's law after the introduction of symbolic notation. Although none of those problems state that the ANS is incorrect, they indicate that some non-trivial problems should be solved to maintain its coherence.

Although we have not been able to find convincing answers to the questions mentioned so far, in the rest of our analysis we still suppose that the two Weber ratios model is correct, and investigate whether the ANS model with two ratios can explain the Indo-Arabic and dots comparisons equally well. This supposition is in line with the different mean error rate of the two notations, and it reflects the views of the proposers of the ANS model (e.g., Piazza et al., 2004; Dehaene, 2007).

ANS Predictions for the Error Rates

In the present section we investigate whether the ANS model predicts the error rate patterns in both notations equally well. We calculated the error rate prediction pattern in our stimulus space for several Weber ratios. Two examples can be seen in **Figure 4**. Weber ratios between 0.05 and 0.25 with a step size of 0.02 were calculated, and fit of the models were calculated for all Weber ratios and for both dot comparison and Indo-Arabic comparison. **Figure 5** shows the R^2 values (right y axes) for the dot comparison and the Indo-Arabic comparison as a function of the Weber ratio (x axis).

First, it is important to clarify that the overall R^2 value difference between the two notations is not appropriate to evaluate the ANS model. While the dot comparison reaches its R^2 maximum at around 0.95, the Indo-Arabic comparison R^2 is not higher than 0.6. The different maximum R^2 values can not only be the result of worse overall fit of the ANS model to the Indo-Arabic comparison, but it can also be the result of the smaller error rate in Indo-Arabic comparison. It is reasonable to suppose that the amount of noise is the same in both notations. However, because of the smaller error rate in Indo-Arabic comparison, the

number pairs related variability is also smaller. Thus, the Indo-Arabic comparison has a lower signal-to-noise ratio. R^2 shows the percentage of the variance the model can explain of the data, but because of the lower signal-to-noise ratio, the percentage of the variance a perfect model could explain is smaller, thus, the maximum R^2 a perfect model could reach is also lower. Although the R^2 should be lower for a less appropriate model, here the variance of the R^2 is directed more strongly by the signal-to-noise ratio. This is another reason why the overall R^2 cannot be used to contrast the model's prediction in the two notations, but a more indirect analysis is required.

Several properties of the ANS model are important, which properties can be used to assess how correct the model is for the two notations. These properties can also show why a more traditional model comparison method is not sufficient.

- (1) **Consistent predicted mean error rates and predicted performance patterns (R^2 values).** Because the ANS model predicts the mean error rate directly, a model with appropriate Weber fraction should find the mean error rate of the measured performance. Additionally, because according to the ANS model the exact shape of the predicted performance (performance pattern) depends on the Weber-fraction of the representation⁷, it also means that a linear fit of that prediction to the measured data should show the highest goodness of fit, when the model uses the appropriate Weber-fraction. Combining these statements, when the appropriate Weber-fraction is found, (a) the model should show the error rate prediction, and at the same time (b) it should show the highest goodness of fit (e.g., highest R^2 value) reflecting that the model finds the shape of the performance across the stimulus space.
- To determine the Weber ratios for the two notations, we looked for the mean error rates of Weber ratios that are equal

⁷In other words, ANS predicted performance patterns with different Weber-fraction, e.g., the two error rates shown in **Figure 4**, cannot be fitted perfectly with a linear transformation.

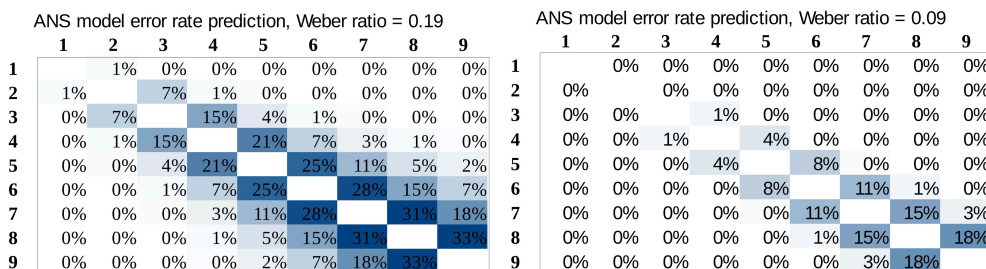


FIGURE 4 | Error rate predictions of the ANS model in our full stimulus space for two Weber ratios. The Weber ratios were determined based on the mean error rates, see **Figure 5** and the text.

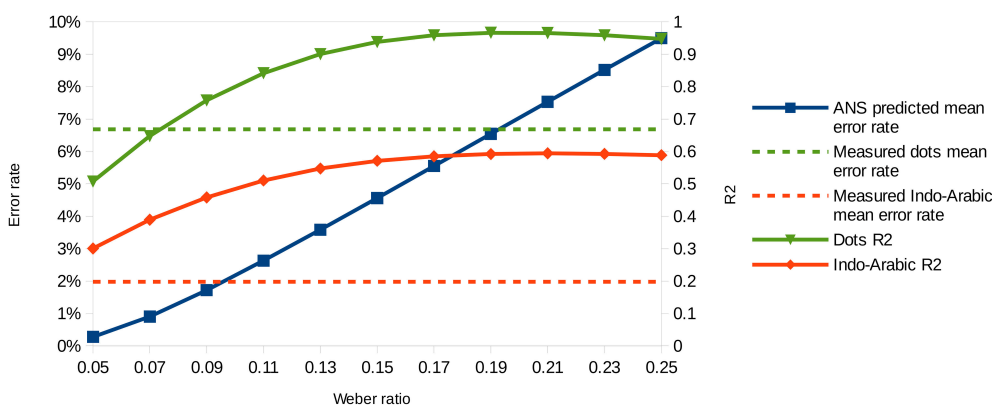


FIGURE 5 | Predicted mean error rates (left y axis) as a function of Weber ratio, and measured mean error rates (left y axis) of the two notations. Goodness of fit (right y axis) as a function of Weber ratio for the dot comparison and the Indo-Arabic comparison.

with the measured mean error rates of the two notations. **Figure 5** shows the predicted mean error rate (left y axis) as a function of Weber ratios (x axis), and the measured Indo-Arabic and dot mean error rates (dashed horizontal lines). Intersections of the prediction (solid line with squares) and the measured data (dashed horizontal lines) specify the Weber ratios of the two notations. According to this, the Weber ratio of the dot comparison should be around 0.19, and the Weber ratio of the Indo-Arabic comparison should be around 0.09. The 0.19 value for non-symbolic stimuli is indeed a typical Weber ratio according to former studies (see for example the results of an extensive measurement in Halberda and Odic, 2014; or the summary of Piazza, 2010 for a review about the development of the Weber ratio). One can note that in the measured data the large ratio cells (e.g., 2 vs. 8, or 10 vs. 45) sometimes show a larger than 0% error rate (**Figure 3**), which is not in line with the prediction of the model (**Figure 4**), reflecting a base error rate, which is independent of the specific number pairs. Because the model cannot account for this error rate which is independent of the comparison stage, it could be more appropriate to subtract this base error rate (around 1%) from the measured error rate (lowering the horizontal dashed line on **Figure 5**). This correction would decrease the Weber ratios by a value around 0.02. All the following results are presented with the 0.19 and 0.09 Weber ratio values, although the same

result patterns could be seen with the corrected 0.17 and 0.07 values, too.

After specifying the Weber ratios of the comparisons for the two notations, one can check if those Weber ratios also show the highest R^2 values. As discussed above, because the goodness of fit should be highest when the Weber ratio is specified correctly (i.e., the model should produce exactly the shape that was measured), the model predicts that the best fit (e.g., the highest R^2) can be obtained with the Weber ratio that is in line with the mean error rate of the notation. With all other Weber ratios the goodness of the fit should be worse. In the dot comparison task the R^2 indeed reaches its maximum around 0.19 Weber ratio, which Weber ratio was predicted based on the measured mean error rate. Thus, the ANS model predicts correctly that the Weber ratio of the best fitting pattern and the Weber ratio based on the mean error rates are approximately the same values. However, in the Indo-Arabic comparison the best R^2 value is around 0.2 Weber ratio, which is much larger than the 0.09 ratio specified with the mean error rate. This suggests that the ANS model cannot predict correctly the shape of the error rate pattern and the mean error rate at the same time in this symbolic comparison.

(2) **Predicted error rate patterns.** Based on the specified Weber ratios we can compare the predicted and the measured error

rate patterns for the whole stimulus space, which can reveal further details how the ANS model prediction deviates from the measured symbolic comparison data. **Figure 4** actually shows the predictions of the model for the Weber ratios with the identified dot and Indo-Arabic Weber ratios, thus, these patterns can be directly compared with the measured data (**Figure 3**). The difference of the measured and the predicted data can be seen in **Figure 6**. Because the model predicts directly the error rates, **Figure 6** can be considered almost as the residuals after fitting the model to the measured data. Positive values show that the model underestimates the measured error rate, while negative values show that the model overestimates the actual error rate. In both notations the model and the actual data show systematic biases, however, they are qualitatively different in nature. (2a) In the dot comparison the misfit of the model is present because the measured data show an asymmetry related to the order of the stimuli, and the model cannot handle this asymmetry. In small ratio pairs large-small number pairs are responded to with smaller error rates (and faster, see **Figure 3**) than small-large number pairs. This effect can be the temporal congruity effect, in which large-small order pairs are handled faster than the small-large order pairs when the instruction is to choose the larger value (Schwarz and Stein, 1998). The effect may appear in our data if participants process the left stimulus first, which is consistent with the Western reading direction. The size of the temporal congruity effect is proportional to the difference of the onset of the two values, and disappears when the two stimuli are presented simultaneously (Schwarz and Stein, 1998). This latter property might explain why in our data the effect is only visible when the processing time is slow. It was proposed that the statistical feature of the data could be used to produce the effect: large numbers have higher probability to be the higher number in a pair, and according to this property, the decision criteria may be modified (Schwarz and Stein, 1998). Otherwise the prediction of the ANS model is relatively correct. (2b) On the other hand, residuals in the Indo-Arabic comparison show a completely different misfit. The model supposes that the error rate is very low for most of the number pairs, and error rate increases steeply for small ratio numbers. Instead of this pattern, measured error rates show that the small ratio number pairs do not show such a high error rate, and error rate starts to increase with larger distance in contrast with the model's prediction. These differences can be seen on the residuals as large overestimation for small ratios, and medium underestimation for medium ratios by the model. (These patterns remain if one would use the base error rate corrected 0.17 and 0.07 Weber ratios, although overall the models would underestimate the measured errors.) These observations suggest that while the ANS model predicts the ratio-based comparison error rates relatively correctly (except the order-based preference for the large-small stimuli in low ratio pairs, which asymmetric effect could be an additional effect), the model cannot describe appropriately the Indo-Arabic comparison error rate pattern.

(3) **Linear regression parameters of the model.** The found parameters of the fitting procedure shed additional light on how the ANS model fails to explain symbolic comparison data. The ANS error function predicts the error rate directly, therefore, with the appropriate Weber ratio the equation of the fit should be $measured_error = 1 \times predicted_error + 0$. How do the parameters change across different Weber values? In the dot comparison task, for example for an incorrectly small 0.07 Weber ratio the fitted function is $2.83 \times model + 0.04$. This high slope is reasonable, because the small Weber ratio predicts too small error rates that should be increased to fit the measured data. For larger Weber ratio the slope gradually decreases, and with the 0.19 Weber ratio (that was specified with the mean error rate) the function is $0.91 \times model + 0.01$, in which the slope is rather close to the expected 1 value that the ANS predicts. In the Indo-Arabic comparison for a 0.07 Weber ratio the estimated function is $0.56 \times model + 0.01$, which is decreasing further as the Weber increases, and for 0.09 Weber ratio the function is $0.37 \times model + 0.01$. These much lower than 1 slopes reflect that the model predicts too sudden increase with small ratios (as observed in the direct comparison of the measured data and the model), and the fit is better when the model is flattened. Again, linear fit of the different Weber ratio models shows that while the ANS predicts correctly the dot comparison error rates, the model cannot predict the Indo-Arabic comparison.

To summarize, in a more extensive analysis, we found that on one hand the ANS model's prediction is coherent in the dot condition: a 0.19 Weber ratio correctly predicts the mean error rate, the relative shape of the error rates and the specific error rates for the number pairs. On the other hand, in the Indo-Arabic comparison the ANS model predicts a too steeply increasing error rate for small ratios, reflected in incoherent fit results. Again, the ANS model proposes that beyond the Weber fraction differences between the two notations, the same error function should hold for both notations (Dehaene, 2007), therefore, the lack of the precise ANS model description of the symbolic comparison is not the consequence of the notations specific processes. Thus, these results contradict the ANS model in its current form that suggests that both symbolic and non-symbolic comparisons are handled by the same type of representations.

Linear Similarity of the Reaction Time Patterns

Group mean of dot comparison time for the whole stimulus space was fit to the group mean of Indo-Arabic comparison time for the whole stimulus space (right of **Figure 3**) According to the result, $Indo-Arabic_RT = 0.17 \times dot_RT + 474.8$, $R^2 = 0.684$. Residuals of the fit (**Figure 7**) show an observable systematic pattern. The fitted dot data underestimate Indo-Arabic reaction time for small distance pairs, and overestimates it for large distance pairs. Additionally, the fitted dot data overestimate the cells with 1 and 9 values, similar to an end effect (see **Figure 2**). To test the presence of these effects in the residuals, multiple linear regression was used with linear distance effect and end

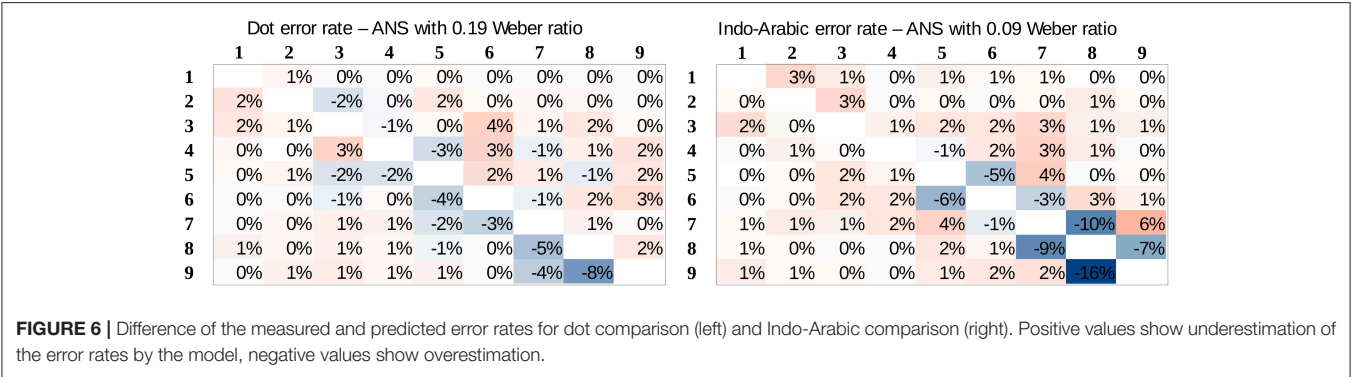


FIGURE 6 | Difference of the measured and predicted error rates for dot comparison (left) and Indo-Arabic comparison (right). Positive values show underestimation of the error rates by the model, negative values show overestimation.

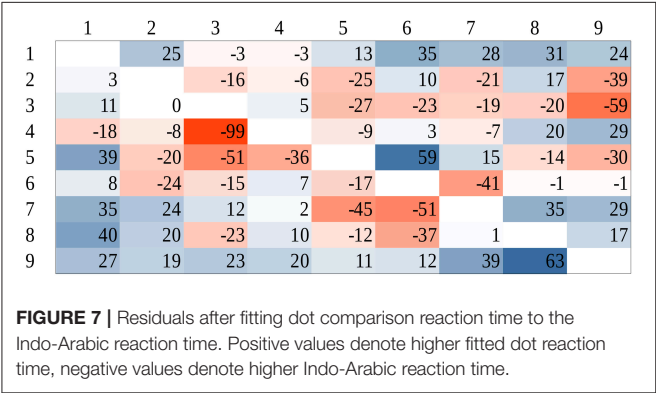


FIGURE 7 | Residuals after fitting dot comparison reaction time to the Indo-Arabic reaction time. Positive values denote higher fitted dot reaction time, negative values denote higher Indo-Arabic reaction time.

effect regressors (see **Figure 2**), and the residual pattern was used as the dependent variable. Only the end effect regressor was significant (slope is 22.3, $p = 0.002$), while the distance effect was not (slope is 1.3, $p = 0.452$). The statistical lack of the distance effect contradicts the observable pattern, although visual inspection could be unreliable. One source of this contradiction could be the insufficient signal-to-noise ratio, and outliers might decrease the statistical power. After excluding two outlier cells (4-3 and 5-6) the correlation between the linear distance effect and the residuals when both numbers are in the 2–8 range (i.e., without the end effect cells) becomes significant, $r(38) = 0.28$, $p = 0.015$.⁸ Thus, because of the observed systematic patterns in the residuals, the reaction time pattern of the dot and Indo-Arabic comparisons cannot be transformed to the other linearly, contrary to the former descriptions.

Although, as we have discussed, this analysis cannot be considered as a sufficiently precise method, it can be used to judge whether this type of reasoning has been cited correctly

⁸One might suggest that the apparent distance effect in the residuals could be the artifact of fitting the dot data to the Indo-Arabic data with the end effect in the Indo-Arabic notation, and with the lack of the effect in the dot notation: because there is a stepwise change at the edge of the Indo-Arabic stimulus space, the “outer end” of the fitted distance effect will be lowered, creating a higher slope in the fitted line and a gradually increasing effect in the residuals (a distance like effect). However, such an artifact should underestimate large distance cells, while our data show an overestimation for those cells. Therefore, the distance effect in the residuals cannot be the artifact of the end effect in the Indo-Arabic notation.

to support the common mechanism behind symbolic and non-symbolic number processing. Our results suggest again that this test cannot confirm that non-symbolic and symbolic numbers are processed by the same system.

Diffusion Model Analysis

The diffusion model analysis can be more sensitive than the error rate analysis, and more appropriate than the reaction time analysis by present-day standards. Drift rates for all number pairs and participants were calculated in both notations. The mean drift rates of the participants for the full stimulus space in the two notations are displayed in **Figure 8**. At first sight it is observable that drift rates show the distance and the size effects in both notations, and the dot comparison is harder than the Indo-Arabic comparison (dot drift rates are smaller), in line with the error rate and the reaction time data.

Drift Rate and Task Difficulty

The values shown in **Figure 8** are displayed in a different way in **Figure 9**. In **Figure 9** drift rates are displayed as the function of the difficulty of the task for the two notations. According to the current theories, the observable function in **Figure 9** could be proportional, $drift_rate = k \times task_difficulty$ (Palmer et al., 2005; Dehaene, 2007), or it could also include a power term as a generalization, $drift_rate = k \times task_difficulty^\beta$, although the exponent is often close to 1, thus the first, proportional model approximates the second, power model. In the ANS model, task difficulty is measured as stimulus strength, which is calculated with the *distance/large_number* function as suggested by Palmer et al. (2005) for psychophysics comparison.⁹ There are different properties that should be seen on this figure for any tasks or for

⁹Dehaene (2007) suggests that the difficulty of the task could be expressed as the logarithm of the ratios of the numbers, although that description is not entirely explicit how this function was found. One possibility is that this function was the one that could offer a linear relation between the difficulty of the task and the drift rates presented in that description. We also tested our data with the log(ratio) task difficulty scale, and the results could not be described neither with the proportional model (the curve is clearly non-linear), nor with the power model (the model strongly overestimates the drift rates for the easy tasks). However, Dehaene (2007) (a) used a more restricted diffusion model parameter recovery method, than the EZ diffusion model (although in the same paper EZ diffusion model was also used, its detailed results were not reported), and (b) he analyzed multi-digit number comparison. These differences can explain why a different expression was found as the measure of the task difficulty.

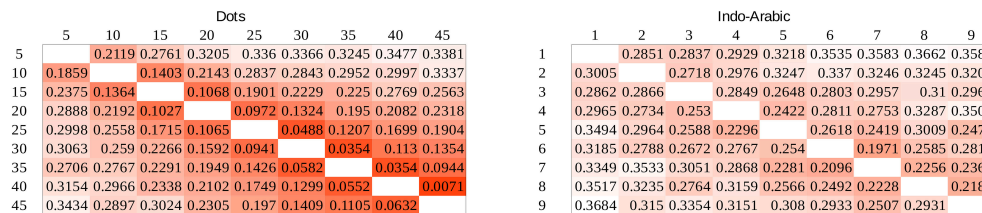


FIGURE 8 | Drift rates in the full stimulus space in dot comparison (left) and in Indo-Arabic comparison (right). Smaller values mean more difficult task.

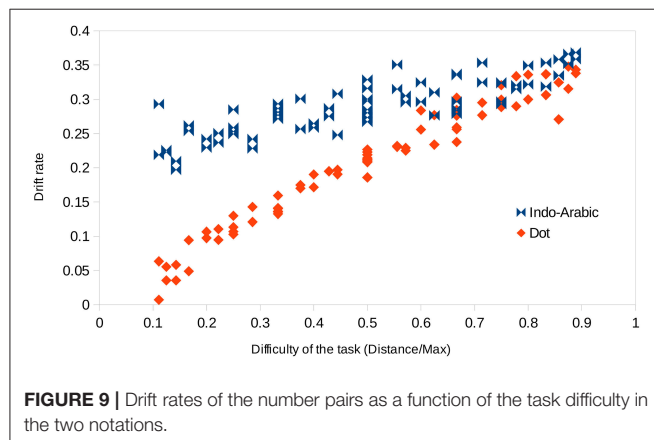


FIGURE 9 | Drift rates of the number pairs as a function of the task difficulty in the two notations.

tasks solved by an analog system. (1) Easier tasks should show higher drift rates, i.e., in **Figure 9** larger values on the x axis should go with larger values on the y axis, showing a positive slope for the curves. This is the case in both notations. However, while in the dot comparison the task difficulty and the drift rate are related more strictly (showing relatively small variance or error around a presumed regression curve), the same relation in the Indo-Arabic notation is much more noisy. (This is not caused by the cells involved in the end effect in Indo-Arabic comparison: after removing those cells, the difference is still visible.) This result is in line with a former study, finding that reaction time is better explained by the ratio in dot comparison task than in Indo-Arabic comparison task (Lyons et al., 2015b p. 1027). This might reflect that while the *distance/large_number* expression suggested by the ANS model might describe the difficulty of the dot comparison relatively well, it might not be applied readily for the Indo-Arabic notation. (2) In an analog representation when the two signals almost completely overlap (i.e., two almost equal properties are shown) the system is hardly able to compare the two properties, which should result in a close to 0 drift rate in the diffusion model (i.e., no evidence is offered for the decision). On **Figure 9** the difficulty is measured as *distance/large_number*, and an indistinguishable pair has a *0/large_number* value, which is 0. Thus, when difficulty tends to zero, drift rate should tend to zero, too, therefore, the intercept of the curves should be zero (Palmer et al., 2005; Dehaene, 2007). This is the case in the dot comparison condition, but Indo-Arabic comparison clearly shows a much higher intercept, somewhere around the 0.2 drift

rate. This 0.2 intercept is in line with another single digit Indo-Arabic comparison task (Krajcsi et al., 2016), and with the non-zero intercept in multi-digit Indo-Arabic comparison (Dehaene, 2007). Again, these results show that while the dot comparison works according to the ANS model, the Indo-Arabic comparison follows other rules.

The 0 intercept of the dot comparison task also confirms that the use of the EZ diffusion model is at least partly appropriate, because its result correctly reflects an important property of an analog mechanism, therefore validating the EZ method.

Dehaene (2007) analyzed a similar data of an Indo-Arabic multi-digit comparison task, and he also found that the intercept of the drift rate function is larger than zero. We note that a multi-digit symbolic comparison might be a multi-step processing (Hinrichs et al., 1982; Poltrock and Schwartz, 1984; Krajcsi and Szabó, 2012), while diffusion model analysis is appropriate only for short, one cycle processing tasks (Wagenmakers et al., 2007), thus, the diffusion model analysis of multi-digit symbolic numbers should be handled cautiously. Still, independent of this problem, it is important to see how these results, which seemingly contradict the ANS model, could be interpreted to support the classic view. To explain the results in the ANS framework, Dehaene (2007) suggested that there could be two subsystems with two different Weber ratios working in a parallel way, and the interaction of these two subsystems could form the higher than zero intercept and the low slope for the Indo-Arabic number comparison. No further explanation was offered how the two subsystems could form this curve. We think that this two subsystems explanation raises some critical issues. First, it is hard to find why the interaction of two systems will produce high drift rate (and high intercept), when both systems can offer only low drift rates, if the stimuli are almost the same. One reasonable combination of the two drift rates could be the addition of the two values, but adding two small values, that are close to zero (as supposed by the ANS model), cannot result in a relatively high 0.2 value. As a more conceptual phrasing, if none of the two subsystems can differentiate between very small differences, why should any combinations of those analog systems perform much better? Another reasonable combination of the two drift rates is that the higher drift rate should be applied, because the less precise subsystem cannot add any extra information to the already more precise subsystem. Again, it is still not clear how the intercept could increase radically. Another problem with this ANS explanation comes from the low slope of the Indo-Arabic drift rate curve. Dehaene (2007) suggests that in the linear

model ($\text{drift_rate} = k \times \text{task_difficulty}$) k is related to the Weber ratio: smaller Weber ratio (higher sensitivity) causes higher slope. Indeed, in the linear model the Weber ratio can be present only in that parameter. Now if we have a k_{dot} slope observed in the dot comparison task, the $k_{\text{Indo-Arabic}}$ slope in the more sensitive Indo-Arabic subsystem should be higher. If those parameters are combined, then again one option is to add the slopes, or another option is to use the larger slope. Both options predict a slope that is larger than the k_{dot} , however, the result shows a smaller value. In a more conceptual rephrase of this problem, the lower slope of the Indo-Arabic comparison suggests a higher (less sensitive) Weber ratio, which contradicts the idea that the Indo-Arabic comparison must be more sensitive than the dot comparison. Overall, we cannot see how the ANS model could explain a drift rate curve with high intercept and low slope, and we propose that the analysis of the Indo-Arabic comparison drift rate data as a function of task difficulty is not in line with the ANS or any other representation working according to Weber's law.

Drift Rate and Representational Overlap

While in the previous analysis the task difficulty was expressed by the relation of the two numbers, one can also incorporate the Weber ratio. The overlap of the representations of the two numbers can be calculated, that depends on the two values and the Weber ratio. The ANS model has another prediction that can be tested here: according to the model, the representational overlap predicts the drift rates in a comparison task. In contrast with the previous task difficulty vs. drift rate analysis, this relation of the drift rates and representational overlap is independent of the notation, because the different Weber ratios of the two notations are already incorporated in the overlap values.

To test whether drift rates depend purely on the representational overlap we calculated the representational overlap for all number pairs in our stimulus space for the two Weber ratios specified earlier. To calculate the overlap of two numbers, two Gaussian distributions were created on a linear scale, with the mean of the two numbers to be compared, and standard deviation was the product of the numbers and the Weber ratio (Halberda and Odic, 2014). Representational overlap values can be seen in **Figure 10**.

Left side of **Figure 11** shows the drift rates as a function of representational overlap in the two notations. In the data for small overlaps the signs of the two notations largely overlap, and to show the potentially hidden dot data, dot data are shifted to the right by 0.01. Also, because the data are hard to explore for small overlap values, the same plot is displayed on a log overlap scale on the right of **Figure 11**. The dot data are not shifted on the latter plot.

According to the ANS model same representational overlap values should result in same drift rate values, independent of the Weber ratio. While for small overlap values the drift rates of the two notations vary in the same range in line with the ANS prediction, for large overlap values Indo-Arabic drift rates are higher than the appropriate dot drift rates, contradicting the ANS model. (This is not caused by the end effect in Indo-Arabic notation: most of the high drift rate values in the large overlap range are not involved in the end effect. Additionally,

the same pattern can be seen with the 0.17 and 0.07 Weber rates which are based on the corrected base error rate.) These data, again, show that the ANS model cannot describe the appropriate representations for both notations.

We also note that while there could be uncertainties whether EZ-diffusion model works correctly, in the current analysis all predictions of the ANS model in the dot comparison task proved to be correct, validating the EZ-diffusion model at the same time. This validation confirms that this simple to use diffusion parameter recovery method can be applied appropriately in the current comparison task.

GENERAL DISCUSSION

The present work investigated whether symbolic Indo-Arabic number comparison and non-symbolic dot comparison can be described by the same model, as predicted by the widely accepted ANS model, or whether the two notations show systematic differences as suggested by the increasing body of evidence and some alternative accounts of symbolic number processing. Although formerly the ANS description for different notation comparisons has been tested, and the fit was found to be satisfactory, the similarity between the ANS and the recently proposed DSS model predictions required a more rigorous and extensive test.

Our results investigating several properties of the ANS model consistently showed that while the ANS model describe several behavioral aspects of the non-symbolic dot comparison relatively well, the symbolic Indo-Arabic comparison deviated from the ANS description at several points. More specifically, (1) while the ANS model predicts the error rate pattern correctly and consistently for non-symbolic dot comparison, it predicts too high error rates in Indo-Arabic comparison for the small ratio pairs, and too low error rates for medium ratio pairs. (2) The reaction time patterns of the two notations have different shapes which cannot be fitted linearly without systematic residuals, although early description of the comparison task reaction time would suggest a stricter similarity between the two patterns. (3a) In the diffusion model framework, while the dot drift rates are more clearly proportional to the difficulty of the task as defined in the ANS model, the relation between the Indo-Arabic drift rates and the ANS derived task difficulty is noisier. (3b) While the dot drift rates tend to zero when the number pairs become indistinguishable, the Indo-Arabic drift rates remain relatively high, contradicting the supposed functioning of a noisy analog representation. (3c) Across the notations, the drift rates do not show the same values depending on the representational overlap as suggested by the ANS model, showing that the two notation comparisons cannot be described by the same mechanism. All of these results show that (a) non-symbolic dot comparison and symbolic Indo-Arabic comparison do not rely purely on the same type of mechanism, and (b) while the ANS model can describe the non-symbolic dot comparison, it cannot describe the symbolic Indo-Arabic notation.

One might wonder whether alternative forms of the ANS model could give an account for our findings, either by modifying

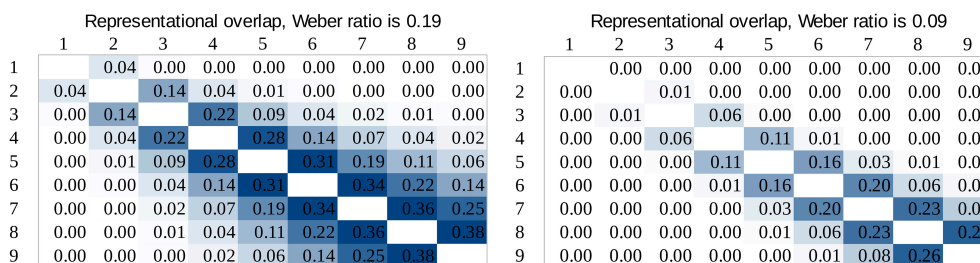


FIGURE 10 | Representational overlap in our stimulus space predicted by the ANS model for Weber ratios 0.19 and 0.9.

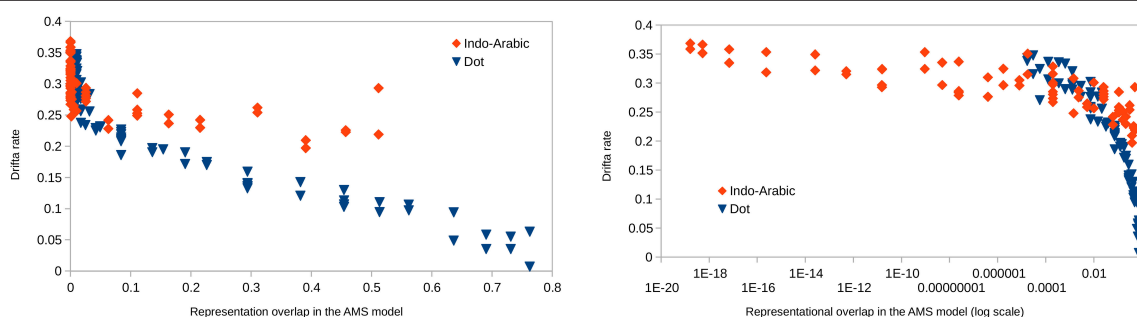


FIGURE 11 | Drift rates as a function of representational overlap in the ANS model in the two notations. Overlap is displayed on linear (left) and logarithmic (right) scale. On the left plot, dot data are shifted by 0.01 to the right not to be covered by the Indo-Arabic data.

the specific functions utilized in the present analyses or by conceptually modifying the model. At least one aspect of our results questions whether this is possible. In Indo-Arabic number comparison the drift rate does not tend to zero when the stimuli become almost indistinguishable, which result cannot be explained by any analog representation working according to the Weber's law. This is an analogous form of the problem that it is difficult to explain how the imprecise ANS could be responsible for precise number processing. If the EZ diffusion model recovered appropriately the drift rates (we indeed found that many properties of the non-symbolic drift rates are in line with the psychophysics model, which validates the EZ model), then the symbolic number comparison cannot be processed by any analog representation working according to the Weber's law, which is a defining feature of the ANS model. Thus, we argue that the ANS model cannot be modified to account for the present findings.

One might also wonder whether shorter presentation of the dot stimuli could modify the results, because that could ensure that the diffusion model analysis handles a single step decision process instead of a multi-step counting process. However, the relatively precise prediction of the ANS model in dot comparison reflects that the current stimuli are successful enough to show the appropriateness of the ANS model, and further refinements can only improve this appropriateness. More generally, because the current design and stimuli were already appropriate to show that the ANS model describes non-symbolic comparison correctly, there is no need to

further improve the current methods using the non-symbolic stimuli.

Beyond the current empirical results, suggesting that only non-symbolic comparison seems to be supported by an analog representation, but not symbolic comparison, we briefly summarize some non-trivial key problems of the ANS model explaining symbolic number processing. (1) As we have mentioned, how could an imprecise system, as the ANS, solve precise symbolic comparison? Even a smaller Weber ratio (more sensitive system) is inappropriate to solve this issue. (2) If a supplementary precise system helps to solve precise symbolic comparison, why is this system invisible in a sense that dominant part of the variance in the comparison performance is purely influenced by the ANS? Additionally, why is the ANS thought to dominantly influence performance in cases when it cannot solve the problem at all? (3) If the supplementary precise system has an effect on the performance, how do we know by looking at the performance that the ANS is also activated in a comparison task? If performance is partly comprised of a hypothetical precise system, then without specifying that precise component, one can not find the rest of the performance that could support the ANS processing either.

To summarize, all of our results show that symbolic and non-symbolic comparisons show several critical differences, and while the ANS model can successfully describe the non-symbolic dot comparison, it cannot account for many features of the symbolic Indo-Arabic comparison. Therefore, we argue that while non-symbolic comparison is supported by the ANS,

symbolic comparison and number processing is supported by an alternative system. Further research can confirm whether the increasing amount of data suggest correctly that symbolic and non-symbolic numbers are processed by different types of systems, and if so, what representation is utilized to process symbolic numbers.

ETHICS STATEMENT

All studies reported here were carried out in accordance with the recommendations of the Department of Cognitive Psychology ethics committee with written informed consent

from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

All authors listed, have made substantial, direct and intellectual contribution to the work, and approved it for publication.

ACKNOWLEDGMENTS

We thank Krisztián Kasos, Ákos Laczkó, and Katalin Oláh for their comments on an earlier version of the manuscript.

REFERENCES

- Balakrishnan, J. D., and Ashby, F. G. (1991). Is subitizing a unique numerical ability? *Percept. Psychophys.* 50, 555–564.
- Bulthé, J., De Smedt, B., and Op de Beeck, H. P. (2014). Format-dependent representations of symbolic and non-symbolic numbers in the human cortex as revealed by multi-voxel pattern analyses. *Neuroimage* 87, 311–322. doi: 10.1016/j.neuroimage.2013.10.049
- Bulthé, J., De Smedt, B., and Op de Beeck, H. P. (2015). Visual number beats abstract numerical magnitude: format-dependent representation of arabic digits and dot patterns in human parietal cortex. *J. Cogn. Neurosci.* 27, 1376–1387. doi: 10.1162/jocn_a_00787
- Crossman, E. R. F. W. (1955). The measurement of discriminability. *Q. J. Exp. Psychol.* 7, 176–195. doi: 10.1080/17470215508416692
- Dakin, S. C., Tibber, M. S., Greenwood, J. A., Kingdom, F. A. A., and Morgan, M. J. (2011). A common visual metric for approximate number and density. *Proc. Natl. Acad. Sci. U.S.A.* 108, 19552–19557. doi: 10.1073/pnas.1113195108
- Damarla, S. R., and Just, M. A. (2013). Decoding the representation of numerical values from brain activation patterns. *Hum. Brain Mapp.* 34, 2624–2634. doi: 10.1002/hbm.22087
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N
- Dehaene, S. (1996). The organization of brain activations in number comparison: event-related potentials and the additive-factors method. *J. Cogn. Neurosci.* 8, 47–68. doi: 10.1162/jocn.1996.8.147
- Dehaene, S. (2007). “Symbols and quantities in parietal cortex: elements of a mathematical theory of number representation and manipulation,” in *Sensorimotor Foundations of Higher Cognition, Vol. XXII*, eds P. Haggard, Y. Rossetti, and M. Kawato (Cambridge: Harvard University Press), 527–574.
- Dehaene, S., and Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition* 43, 1–29. doi: 10.1016/0010-0277(92)90030-L
- Eger, E., Sterzer, P., Russ, M. O., Giraud, A.-L., and Kleinschmidt, A. (2003). A supramodal representation in human intraparietal cortex. *Neuron* 37, 719–725. doi: 10.1016/S0896-6273(03)00036-9
- Gebuis, T., and Reynvoet, B. (2012). Continuous visual properties explain neural responses to non-symbolic number. *Psychophysiology* 49, 1649–1659. doi: 10.1111/j.1469-8986.2012.01461.x
- Halberda, J., and Odic, D. (2014). “The precision and internal confidence of our approximate number thoughts,” in *Evolutionary Origins and Early Development of Number Processing*, eds D. C. Geary, D. B. Berch, and K. Mann Koepke (Waltham, MA: Academic Press), 305–333.
- Halberda, J., Mazzocco, M. M. M., and Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature* 455, 665–668. doi: 10.1038/nature07246
- Hinrichs, J. V., Berie, J. L., and Mosell, M. K. (1982). Place information in multidigit number comparison. *Mem. Cognit.* 10, 487–495. doi: 10.3758/BF03197652
- Holloway, I. D., and Ansari, D. (2009). Mapping numerical magnitudes onto symbols: the numerical distance effect and individual differences in children’s mathematics achievement. *J. Exp. Child Psychol.* 103, 17–29. doi: 10.1016/j.jecp.2008.04.001
- Huber, S., Nuerk, H.-C., Willmes, K., and Moeller, K. (2016). A general model framework for multisymbol number comparison. *Psychol. Rev.* 123, 667–695. doi: 10.1037/rev0000040
- Kallai, A. Y., and Tzelgov, J. (2009). A generalized fraction: an entity smaller than one on the mental number line. *J. Exp. Psychol. Hum. Percept. Perform.* 35, 1845–1864. doi: 10.1037/a0016892
- Kamienkowski, J. E., Pashler, H., Dehaene, S., and Sigman, M. (2011). Effects of practice on task architecture: combined evidence from interference experiments and random-walk models of decision making. *Cognition* 119, 81–95. doi: 10.1016/j.cognition.2010.12.010
- Kaufman, E. L., Lord, M. W., Reese, T. W., and Volkman, J. (1949). The discrimination of visual number. *Am. J. Psychol.* 62, 498–525. doi: 10.2307/1418556
- Kingdom, F. A. A., and Prins, N. (2010). *Psychophysics: a Practical Introduction*. London: Academic.
- Krajcsi, A. (2017). Numerical distance and size effects dissociate in Indo-Arabic number comparison. *Psychon. Bull. Rev.* 24, 927–934. doi: 10.3758/s13423-016-1175-6
- Krajcsi, A., and Szabó, E. (2012). The role of number notation: sign-value notation number processing is easier than place-value. *Front. Psychol.* 3:463. doi: 10.3389/fpsyg.2012.00463
- Krajcsi, A., Lengyel, G., and Kojouharova, P. (2016). The source of the symbolic numerical distance and size effects. *Front. Psychol.* 7:1795. doi: 10.3389/fpsyg.2016.01795
- Krajcsi, A., Szabó, E., and Mórocz, I. Á. (2013). Subitizing is sensitive to the arrangement of objects. *Exp. Psychol.* 60, 227–234. doi: 10.1027/1618-3169/a000191
- Leibovich, T., and Ansari, D. (2016). The symbol-grounding problem in numerical cognition: a review of theory, evidence, and outstanding questions. *Can. J. Exp. Psychol.* 70, 12–23. doi: 10.1037/cep0000070
- Leth-Steensen, C., Lucas, J., and Petrusic, W. M. (2011). “Modelling SNARC by using polarity codes to adjust drift rates,” in *Proceedings of the 27th Annual Meeting of the International Society for Psychophysics (Herzliya)*, 357–362.
- Lyons, I. M., Ansari, D., and Beilock, S. L. (2015a). Qualitatively different coding of symbolic and non-symbolic numbers in the human brain. *Hum. Brain Mapp.* 36, 475–488. doi: 10.1002/hbm.22641
- Lyons, I. M., Nuerk, H.-C., and Ansari, D. (2015b). Rethinking the implications of numerical ratio effects for understanding the development of representational precision and numerical processing across formats. *J. Exp. Psychol. Gen.* 144, 1021–1035. doi: 10.1037/xge0000094
- Mandler, G., and Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *J. Exp. Psychol. Gen.* 111, 1–22. doi: 10.1037/0096-3445.111.1.1
- Moyer, R. S., and Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature* 215, 1519–1520.
- Nieder, A. (2005). Counting on neurons: the neurobiology of numerical competence. *Nat. Rev. Neurosci.* 6, 177–190. doi: 10.1038/nrn1626

- Noël, M.-P., and Rousselle, L. (2011). Developmental changes in the profiles of dyscalculia: an explanation based on a double exact-and-approximate number representation model. *Front. Hum. Neurosci.* 5:165. doi: 10.3389/fnhum.2011.00165
- Palmer, J., Huk, A. C., and Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J. Vis.* 5, 1–1. doi: 10.1167/5.5.1
- Park, J., and Brannon, E. M. (2013). Training the approximate number system improves math proficiency. *Psychol. Sci.* 24, 2013–2019. doi: 10.1177/0956797613482944
- Patro, K., Nuerk, H.-C., Cress, U., and Haman, M. (2014). How number-space relationships are assessed before formal schooling: a taxonomy proposal. *Dev. Psychol.* 5:419. doi: 10.3389/fpsyg.2014.00419
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends Cogn. Sci.* 14, 542–551. doi: 10.1016/j.tics.2010.09.008
- Piazza, M., Izard, V., Pinel, P., Le Bihan, D., and Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron* 44, 547–555. doi: 10.1016/j.neuron.2004.10.014
- Piazza, M., Mechelli, A., Butterworth, B., and Price, C. J. (2002). Are subitizing and counting implemented as separate or functionally overlapping processes? *Neuroimage* 15, 435–446. doi: 10.1006/nimg.2001.0980
- Pinhas, M., and Tzelgov, J. (2012). Expanding on the mental number line: zero is perceived as the “smallest.” *J. Exp. Psychol. Learn. Mem. Cognit.* 38, 1187–1205. doi: 10.1037/a0027390
- Pollock, S. E., and Schwartz, D. R. (1984). Comparative judgments of multidigit numbers. *J. Exp. Psychol. Learn. Mem. Cognit.* 10, 32–45. doi: 10.1037/0278-7393.10.1.32
- Proctor, R. W., and Cho, Y. S. (2006). Polarity correspondence: a general principle for performance of speeded binary classification tasks. *Psychol. Bull.* 132, 416–442. doi: 10.1037/0033-2909.132.3.416
- Ratcliff, R., and McKoon, G. (2008). The diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* 20, 873–922. doi: 10.1162/neco.2008.12-06-420
- Ratcliff, R., and Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: approaches to dealing with contaminant reaction times and parameter variability. *Psychon. Bull. Rev.* 9, 438–481. doi: 10.3758/BF03196302
- Revin, S. K., Piazza, M., Izard, V., Cohen, L., and Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychol. Sci.* 19, 607–614. doi: 10.1111/j.1467-9280.2008.02130.x
- Rouder, J. N., Lu, J., Speckman, P., Sun, D., and Jiang, Y. (2005). A hierarchical model for estimating response time distributions. *Psychon. Bull. Rev.* 12, 195–223. doi: 10.3758/BF03257252
- Sasanguie, D., Defever, E., Maertens, B., and Reynvoet, B. (2014). The approximate number system is not predictive for symbolic number processing in kindergarteners. *Q. J. Exp. Psychol.* 67, 271–280. doi: 10.1080/17470218.2013.803581
- Sathian, K., Simon, T. J., Peterson, S., Patel, G. A., Hoffman, J. M., and Grafton, S. T. (1999). Neural evidence linking visual object enumeration and attention. *J. Cogn. Neurosci.* 11, 36–51. doi: 10.1162/089892999563238
- Schneider, M., Beeres, K., Coban, L., Merz, S., Susan Schmidt, S., Stricker, J., et al. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: a meta-analysis. *Dev. Sci.* 20:e12372. doi: 10.1111/desc.12372
- Scholz, K. W., and Potts, G. R. (1974). Cognitive processing of linear orderings. *J. Exp. Psychol.* 102, 323–326. doi: 10.1037/h0035873
- Schwarz, W., and Stein, F. (1998). On the temporal dynamics of digit comparison processes. *J. Exp. Psychol. Learn. Mem. Cogn.* 24, 1275–1293. doi: 10.1037/0278-7393.24.5.1275
- Smith, P. L., and Ratcliff, R. (2004). Psychology and neurobiology of simple decisions. *Trends Neurosci.* 27, 161–168. doi: 10.1016/j.tins.2004.01.006
- Stoianov, I., and Zorzi, M. (2012). Emergence of a “visual number sense” in hierarchical generative models. *Nat. Neurosci.* 15, 194–196. doi: 10.1038/nn.2996
- Tzelgov, J., Ganor-Stern, D., and Maymon-Schreiber, K. (2009). The representation of negative numbers: exploring the effects of mode of processing and notation. *Q. J. Exp. Psychol.* 62, 605–624. doi: 10.1080/17470210802034751
- Verguts, T., and Fias, W. (2004). Representation of number in animals and humans: a neural model. *J. Cogn. Neurosci.* 16, 1493–1504. doi: 10.1162/0898929042568497
- Verguts, T., and Van Opstal, F. (2014). A delta-rule model of numerical and non-numerical order processing. *J. Exp. Psychol. Hum. Percept. Perform.* 40, 1092–1012. doi: 10.1037/a0035114
- Verguts, T., Fias, W., and Stevens, M. (2005). A model of exact small-number representation. *Psychon. Bull. Rev.* 12, 66–80. doi: 10.3758/BF03196349
- Vigliocco, G., Vinson, D. P., Damian, M. F., and Levelt, W. (2002). Semantic distance effects on object and action naming. *Cognition* 85, B61–B69. doi: 10.1016/S0010-0277(02)00107-5
- Wagenmakers, E.-J., van der Maas, H. L. J., and Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychon. Bull. Rev.* 14, 3–22. doi: 10.3758/BF03194023
- Welford, A. T. (1960). The measurement of sensory-motor performance: survey and reappraisal of twelve years' progress. *Ergonomics*, 3, 189–230. doi: 10.1080/00140136008930484

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Krajcsi, Lengyel and Kojouharova. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Task Constraints Affect Mapping From Approximate Number System Estimates to Symbolic Numbers

Dana L. Chesney^{1*} and Percival G. Matthews²

¹ Department of Psychology, St. John's University, Jamaica, NY, United States, ² Department of Educational Psychology, University of Wisconsin-Madison, Madison, WI, United States

OPEN ACCESS

Edited by:

Marcus Lindskog,
Uppsala University, Sweden

Reviewed by:

Bert Reynvoet,
KU Leuven, Belgium
Evelyn Kroesbergen,
Radboud University Nijmegen,
Netherlands

*Correspondence:

Dana L. Chesney
chesneyd@stjohns.edu;
dlchesney@gmail.com

Specialty section:

This article was submitted to
Cognition,
a section of the journal
Frontiers in Psychology

Received: 14 May 2018

Accepted: 05 September 2018

Published: 16 October 2018

Citation:

Chesney DL and Matthews PG (2018)
Task Constraints Affect Mapping
From Approximate Number System
Estimates to Symbolic Numbers.
Front. Psychol. 9:1801.
doi: 10.3389/fpsyg.2018.01801

The Approximate Number System (ANS) allows individuals to assess nonsymbolic numerical magnitudes (e.g., the number of apples on a tree) without counting. Several prominent theories posit that human understanding of symbolic numbers is based – at least in part – on mapping number symbols (e.g., 14) to their ANS-processed nonsymbolic analogs. Number-line estimation – where participants place numerical values on a bounded number-line – has become a key task used in research on this mapping. However, some research suggests that such number-line estimation tasks are actually proportion judgment tasks, as number-line estimation requires people to estimate the magnitude of the to-be-placed value, relative to set upper and lower endpoints, and thus do not so directly reflect magnitude representations. Here, we extend this work, assessing performance on nonsymbolic tasks that should more directly interface with the ANS. We compared adults' ($n = 31$) performance when placing nonsymbolic numerosities (dot arrays) on number-lines to their performance with the same stimuli on two other tasks: Free estimation tasks where participants simply estimate the cardinality of dot arrays, and ratio estimation tasks where participants estimate the ratio instantiated by a pair of arrays. We found that performance on these tasks was quite different, with number-line and ratio estimation tasks failing to show classic psychophysical error patterns of scalar variability seen in the free estimation task. We conclude the constraints of tasks using stimuli that access the ANS lead to considerably different mapping performance and that these differences must be accounted for when evaluating theories of numerical cognition. Additionally, participants showed typical underestimation patterns in the free estimation task, but were quite accurate on the ratio task. We discuss potential implications of these findings for theories regarding the mapping between ANS magnitudes and symbolic numbers.

Keywords: approximate number system, symbolic number mapping, number-lines, ratios, estimation

INTRODUCTION

Humans and many nonhuman animals are equipped with a phylogenetically ancient approximate number system (ANS) that allows them to rapidly enumerate the items in a set without counting (Kaufman et al., 1949; Mechner, 1958; Meck and Church, 1983; Feigenson et al., 2004; Izard and Dehaene, 2008). These findings have led many to conclude that the meanings of symbolic numbers

are grounded in part by mapping number symbols (e.g., 5) to their nonsymbolic analogs (e.g., an array of 5 dots) (Nieder and Dehaene, 2009). This obvious symbol-to-referent match is a large part of the appeal of the analog portion of Dehaene's (1992) triple code model and of Piazza's (2010) hypothesis about the ANS' role as a neurocognitive start-up tool for number concepts. Although there is substantial disagreement surrounding ANS-as-foundation arguments (e.g., Lyons et al., 2012; De Smedt et al., 2013; Reynvoet and Sasanguie, 2016; Leibovich et al., 2017; Núñez, 2017), this point of view remains widespread.

Number-line estimation – in which participants place numerical values on a bounded number-line – has become a key task used in research on the link between symbolic numbers and numerical magnitudes (Siegler and Opfer, 2003; Whyte and Bull, 2008; Schley and Peters, 2014). Some consider the spacing and precision of number-line placements to directly reflect the spacing and precision of the magnitudes mapped to symbolic numbers (Siegler and Opfer, 2003; Whyte and Bull, 2008). However, this interpretation of number-line performance remains contested. Some researchers (e.g., Barth and Paladino, 2011) argue that number-line tasks are *proportion judgment* tasks as they require people to estimate the magnitudes of the stimuli relative to the endpoints. Prior research indicates such anchored tasks are fundamentally different from tasks for which participants are free to give any response (Banks and Coleman, 1981; Hollands and Dyre, 2000). As such, task demands may influence participants' mapping responses.

Moreover, there is reason to question the underlying assumption that people can exploit a 1-to-1 map from symbols to their analog numerosities. More than 75 years of research suggest that the vast majority of educated humans cannot accurately make such mappings (Taves, 1941; Kaufman et al., 1949; Indow and Ida, 1977; Krueger, 1984; Izard and Dehaene, 2008; Crollen et al., 2011). In study after study, ANS-based estimations yield under-estimations, and performance varies considerably between participants (Indow and Ida, 1977; Krueger, 1984; Izard and Dehaene, 2008). Given that ANS-based estimation is both inaccurate generally and inconsistent among individuals, it is difficult to see how such a system can be used for grounding symbolic numbers.

Here we seek to clarify principles governing the potential links between ANS-perceived magnitudes and symbolic numbers and how responses based on those links are affected by different task constraints. We investigated how three separate tasks that employ the same sorts of ANS stimuli lead to differences in mapping performance: free estimation, number-line estimation, and ratio estimation.

Predictions

Free Estimation

In free estimation tasks, participants are instructed to give numerical estimates for a range of stimuli whose magnitudes vary on a given dimension, with no given upper bound. This sort of estimation with numerosities has often been described as representing subjective numerical magnitudes in a

logarithmic fashion, such that the perceived distance between stimuli is proportional to the logarithm of the ratio between them (e.g., Moyer and Landauer, 1967; Dehaene, 1992). Hence, the perceived difference between 10 and 20 dots is the same as that between 22 and 44, or that between 32 and 64 dots. Izard and Dehaene (2008) offered a model whereby idiosyncrasies in mapping between logarithmically encoded perceived magnitude and actual symbolic numerical responses results in performance that is typically fit by power functions (e.g., Stevens, 1957; Crollen et al., 2011; but see Cordes et al., 2001, for a linear interpretation). Indeed, performance patterns on such unbounded estimations in general – whether involving numerosities or other magnitudes like auditory volume or light intensities – are typically fit by accelerating or decelerating power functions [perceived stimulus intensity = $C * (\text{Actual stimulus intensity})^B$, where B is the Stevens' exponent e.g., Stevens, 1957; Indow and Ida, 1977; Krueger, 1984; Crollen et al., 2011].

In the ANS-based free estimation task we use here, participants were asked to provide estimates of the numerosity of nonsymbolic numerical stimuli (dot arrays). We expected unbounded estimation with dot arrays to be characterized by compressive power functions (i.e., Stevens' exponent < 1), as is consistent with established theory and prior empirical findings (e.g., Stevens, 1957; Crollen et al., 2011). We also expected estimates to exhibit scalar variability (Cordes et al., 2001; Izard and Dehaene, 2008; Crollen et al., 2011). That is, we expected the variability of estimates to increase in proportion to the size of the stimulus, resulting in a constant coefficient of variation (Whalen et al., 1999; Gallistel and Gelman, 2000; Izard and Dehaene, 2008).

Number-Line Estimation

Our predictions for number line estimation are based on Barth and Paladino's (2011) argument that these tasks cannot properly be categorized as free numerical estimation tasks and that they are actually a form of a proportion judgment task. Number line estimation requires that people estimate the magnitude of one stimulus, the to-be-placed value, relative to two other stimuli, the upper and lower endpoints (Spence, 1990; Hollands and Dyre, 2000; Hollands et al., 2002; but see Opfer et al., 2011). For example, when placing 25 on a 0–100 line (whether symbolic or nonsymbolic), it should be 25 units away from 0, and 75 units away from 100. It should therefore be placed at a point corresponding to the proportion between the stimulus and the sum of the stimulus and its complement ($25/(25 + 75)$), or one fourth of the total length of the line away from 0. No matter what number is estimated, the line must, similarly, be broken into two sections with a constant sum, resulting in a proportion. Spence (1990) offered a cyclical correction to the power model used to describe free estimation that could account for the proportional nature of tasks like number line estimation. This cyclical power model predicts nearly linear performance on number line estimation tasks even given compressive underlying subjective representations of numerical magnitudes (see also Hollands and Dyre, 2000; Hollands et al., 2002; Barth and Paladino, 2011). However,

in approaching linearity, cyclical power models show specific patterns of over- and under-estimation for estimates in different segments of the range defined by specific cut points (see **Figure 1**).

Here, we used an ANS-based number-line estimation task. Participants were instructed to estimate the appropriate placement of a nonsymbolic numerical stimulus (a dot array) on a line segment bounded by nonsymbolic numerical anchors at each end. To date, relatively few studies have attempted to use number-line style tasks using nonsymbolic numerosity (dot arrays) in place of symbolic numbers (Anobile et al., 2012; Sasanguie and Reynvoet, 2013; Kim and Opfer, 2015). None of these investigated whether line estimation with dot array stimuli bears signatures of the cyclical power model as might be predicted following Spence (1990) or Hollands and Dyre (2000). We predicted that these tasks would be fit by a cyclical power model and its characteristics: (a) median estimates should be close to the correct value of the stimulus, (b) the standard deviations of the estimates would *not* show scalar variability patterns, but rather would decrease at both end-point anchors and at the midpoint of the line, and (c) participant responses should exhibit a cyclical pattern of over and then under estimation.

Ratio Estimation

Here, we used an ANS-based ratio estimation task, asking participants to estimate the ratios instantiated by a pair of nonsymbolic numerical stimuli. Recent research suggests that humans and other animals possess a nonsymbolic ratio processing system (RPS) that is tuned to the magnitudes of nonsymbolically instantiated ratios (Jacob et al., 2012; Matthews and Chesney, 2015; Matthews and Lewis, 2016; Matthews et al., 2016; Bonn and Cantlon, 2017).

Unlike proportion judgment tasks, which are typically conceived of as involving judgment of one portion of the whole relative to the judgment of that portion and its complement (Spence, 1990; Hollands and Dyre, 2000; Hollands et al., 2002; Barth and Paladino, 2011), the part:part ratios used in ratio estimation don't have the same constraints. Because the physical magnitudes instantiating the high and low anchors vary considerably from trial to trial, the figure-plus-complement logic of the cyclical power model no longer applies. Accordingly, ratio estimation is posited to proceed from a more direct perceptual mechanism (Jacob and Nieder, 2009; Matthews and Chesney, 2015; Lewis et al., 2016) as opposed to the strategy-bound method that results in cyclical performance on line-based proportion judgment tasks (Spence, 1990; Barth and Paladino, 2011; Cohen and Blanc-Goldhammer, 2011). Indeed, single-cell recordings from primates suggest that there are neurons that respond specifically to visuospatially constructed ratios as opposed to the magnitude of either component of a given ratio (Vallentin and Nieder, 2008).

RPS theories posit that humans can extract the magnitudes of ratios made from a variety of different stimuli, and several studies have directly investigated the human ability to process ratios composed of dot arrays (McCrink and Wynn, 2007; Fabbri et al., 2012; Matthews and Chesney, 2015). Past research

on direct estimation of nonsymbolic ratios made from dot arrays guide our predictions. For instance, Varey et al. (1990), found approximately linear responses in a task similar to our ratio estimation task. Moreover, when Matthews and Chesney (2015) had participants compare symbolic ratios to nonsymbolic ratios, results indicated that participants mapped nonsymbolic dot ratios to numerical ratios in a linear fashion, albeit with a bias that somewhat inflated the size of the nonsymbolic ratios by a constant factor. Finally, in an unpublished pilot study we conducted, we also found that participants' average estimates were largely accurate. These behavioral findings have been complemented by single-cell recordings from primates suggesting that there are neurons that respond specifically to visuospatially constructed ratios as opposed to the magnitude of either component of a given ratio (Vallentin and Nieder, 2008).

Thus, we expected a linear relation between participant estimates and actual stimulus values for ratio estimation tasks (as opposed to the curvilinear relations predicted for free estimation and line estimation tasks). Although we also expected the number-line estimation task to yield roughly linear estimates, we expected those results to diverge from ratio estimates. This is because we expected ratio estimation to proceed from a more direct perceptual mechanism (Jacob and Nieder, 2009; Matthews and Chesney, 2015; Lewis et al., 2016) as opposed to the strategy-bound method that results in cyclical performance on line-based proportion judgment tasks (Spence, 1990; Barth and Paladino, 2011; Cohen and Blanc-Goldhammer, 2011). As result, we did not expect to see such strategy-based cyclical bias patterns with the ratio estimation task.

MATERIALS AND METHODS

Participants

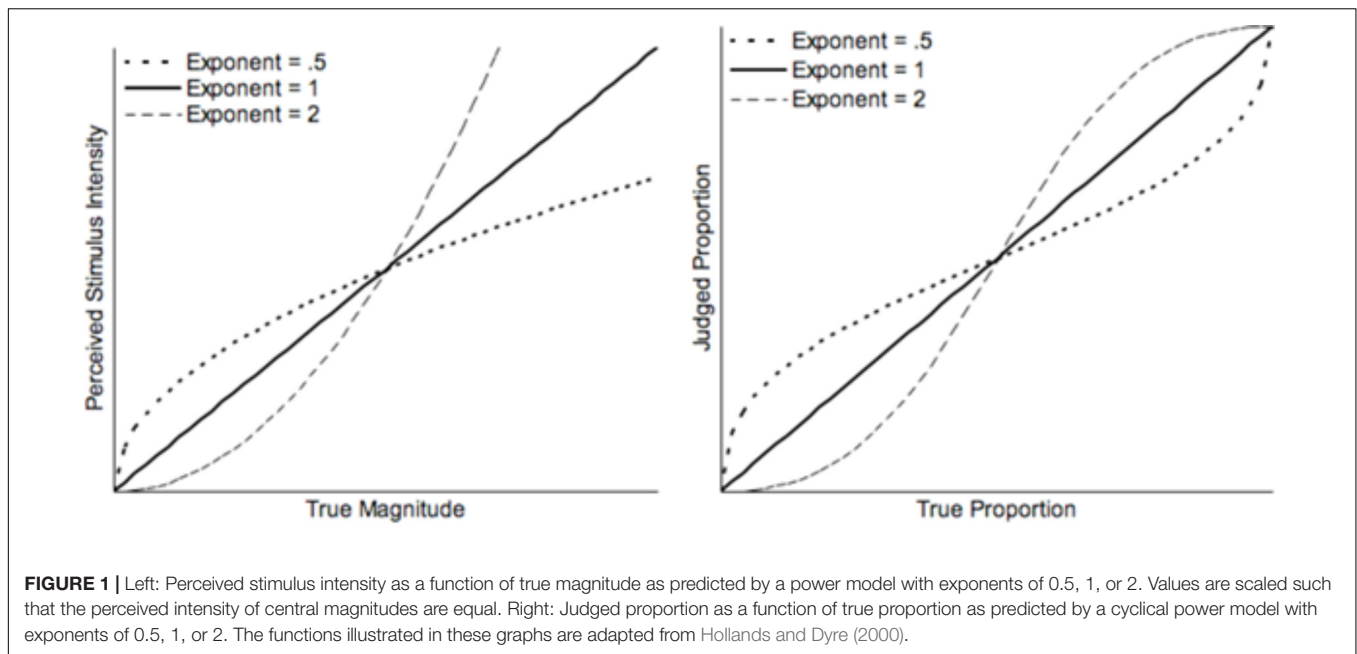
Participants were 31 undergraduates (16 female, 26 white, mean age 19.3 years ($SD = 1.1$ years) at a highly selective, private university in the Midwestern United States who participated for course credit in the Psychology Department.

Materials and Design

All training and testing stimuli were presented using Superlab 4 software (Cedrus Corporation, 2007) on Apple® iMac 5.1 computers running OS10.6. Each computer had a 17" LCD display with a resolution of 1,440 × 900 pixels and a refresh rate of 60 Hz. These screen dimensions subtended approximately 34° × 22° of visual angle with participants seated ~60 cm from the screen. Degrees of visual angle are only approximate as no restraints were used to restrict head motion.

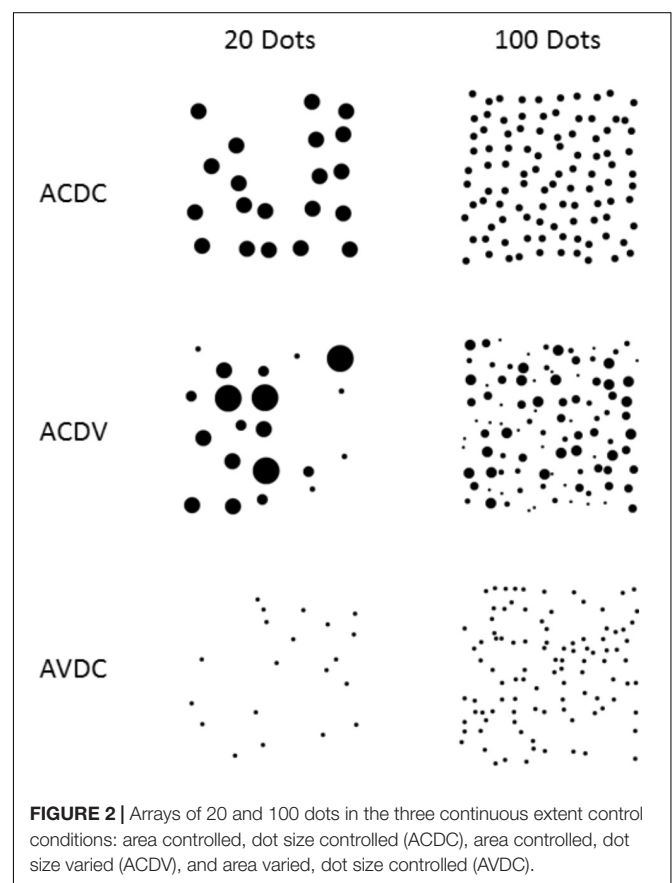
Dot Array Stimuli

Arrays were composed of black dots on a white background. For each array, dot sizes ranged from 1.3 mm to 9.9 mm in diameter (0.1–0.9°), and the minimum distance between dots was 1 mm (0.1°). Dots were arranged randomly in a 76 × 76 mm (7° × 7°) area, such that all arrays had the same convex hull. It was essential to our design that participants used the ANS to estimate the cardinality of the dot arrays, rather than relying



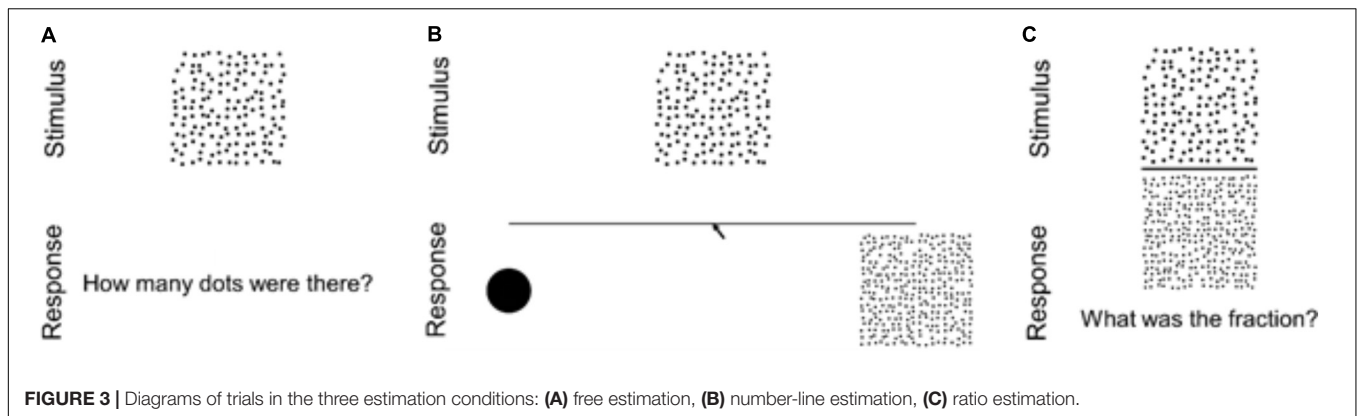
upon counting. Accordingly, the smallest numerosity displayed in a given array was 20 to ensure that other fast enumeration techniques, such as subitizing, could not be employed (see Kaufman et al., 1949; Revkin et al., 2008). The dot arrays in each task ranged in numerosity from 20 to 300 dots. The 17 magnitudes represented were: 20, 40, 60, 80, 100, 120, 140, 150, 160, 180, 200, 220, 240, 260, 280, 290, and 300. Stimuli were presented only briefly (1,500 ms). Brief presentation times have been used successfully to suppress counting in previous work (e.g., Revkin et al., 2008).

To ensure that nonnumeric features of the arrays would not be consistently related to numerosity, we created three different stimuli for each numerosity, with different controls for individual dot size and summed area (see **Figure 2**). In the *area controlled, dot sizes controlled* (ACDC) arrays, the total surface area was controlled such that all arrays had the same total surface area regardless of dot numerosity, and all dots within any given array were of the same size. As a result, the sizes of individual dots in an array varied inversely and density varied directly with the numerosity of the array. In the *area controlled, dot size varied* arrays (ACDV), total surface area was controlled so that all arrays had the same total surface area regardless of dot numerosity. However, individual dot size varied both within and between arrays, such that the size of a given dot did not precisely correlate with array numerosity. As a result, for these arrays, neither total area nor individual dot size was correlated with numerosity (though the mean dot size of an array was inversely correlated with numerosity). In the *area varied, dot size controlled* (AVDC) arrays, all dots were the same size, regardless of the numerosity. As a result, surface area and density increased linearly with the total numerosity of dots presented. These controls mirror those that have been used in previous studies of numerosity perception (Xu et al., 2005; Hurewitz et al., 2006).



Procedure

Participants first completed the ratio estimation block, followed by the number-line estimation block, and finally the free



estimation block (see **Figure 3**). We placed blocks in this order to minimize the likelihood that any block would affect estimation on the subsequent block. Each block began with a set of instructions, using example stimuli that were different from the experimental stimuli. Participants were told that the dot arrays would be presented too quickly for them to count, and that they should “just try to feel out how many dots there are instead of applying a formula.” In all trials, participants pressed a space bar to initiate the trial, then stimuli were briefly presented (1,500 ms), and finally participants were asked to make their responses. If participants did not answer within 15,000 ms, the trial ended automatically. Trial order was randomized within each block. Participants also completed similar tasks involving circle areas, a symbolic number line task, and several mathematics assessments not discussed in this manuscript. We note that, due to experimenter error, one participant completed nearly double the number of trials for each task.

Free Estimation

For each trial, a stimulus array was presented for 1,500 ms immediately after the participants initiated the trial. Once the stimulus disappeared, a textbox appeared asking, “How many dots were there?” Participants entered their answers into a text box via keyboard. After responding, they were prompted to hit return to move on to the next trial. Participants completed 51 trials, one for each of the 17 dot numerosities presented in each of the 3 dot array types.

Number-Line Estimation

For each trial, participants were shown a “number-line” anchored by one dot on the left and 300 dots on the right. Participants were never told the number of dots on the high anchor. When participants hit the space bar to initiate each trial, the line and anchors appeared. After 1,000 ms elapsed, the stimulus array was presented 25 mm above the center of the line for 1,500 ms. Once the stimulus disappeared, participants used a mouse to indicate the position on the line corresponding to the stimulus numerosity. The line and anchors remained on the screen throughout the duration of each trial. After responding, they were prompted to hit return to move on to the next trial. Participants completed 51 trials, one for each of the 17 dot numerosities presented in each of the 3 dot array types.

Ratio Estimation

In ratio estimation trials, participants were instructed to estimate the ratio between the numbers of dots in the two arrays composing each stimulus. Each stimulus was presented for 1,500 ms immediately after the participants initiated a trial. Once the stimulus disappeared, a textbox appeared asking, “What was the fraction?” Participants then typed their answers into a text box via the keyboard. After responding, they were prompted to hit return to move on to the next trial. Participants completed 51 trials, one for each of the 17 dot numerosities in each of the three formats used in the free estimation and number-line estimation blocks, with the 300 dot stimulus of the matching ACDC, ACDV, or AVDC type in the denominator position (e.g., 20 dots/300 dots, 150 dots/300 dots). Additional trials using denominators of other numerosities were also included, however, only the 300 denominator trials are presented in the results here, so as to increase comparability between blocks.

RESULTS

Coding

On the free estimation trials, analyses used participants’ raw responses. One outlier (“9101”) was dropped from consideration. Participants’ spatial position responses on the number-line estimation trials were converted to numerical form corresponding to each response’s relative location on a 1–300 linear number-line. For example, a click on the midpoint of the line was coded as a response of 150. Responses on the ratio estimation trials were first converted to decimal format (e.g., $\frac{1}{2}$, $\frac{50}{100}$, and $\frac{150}{300}$ were all coded as 0.5). Decimal answers (e.g., 0.8) were also accepted. Trials where participants failed to provide a complete ratio (19 trials) or provided values greater than $\frac{5}{2}$ (5 trials) were dropped from consideration. Coded values were then multiple by 300 to place them on the same scale as the Free estimation and number-line estimation tasks for the purposes of analysis.

Analysis

For each of the 51 stimuli (the 17 magnitudes in the three format) in each of the three blocks, we found the participants’ median

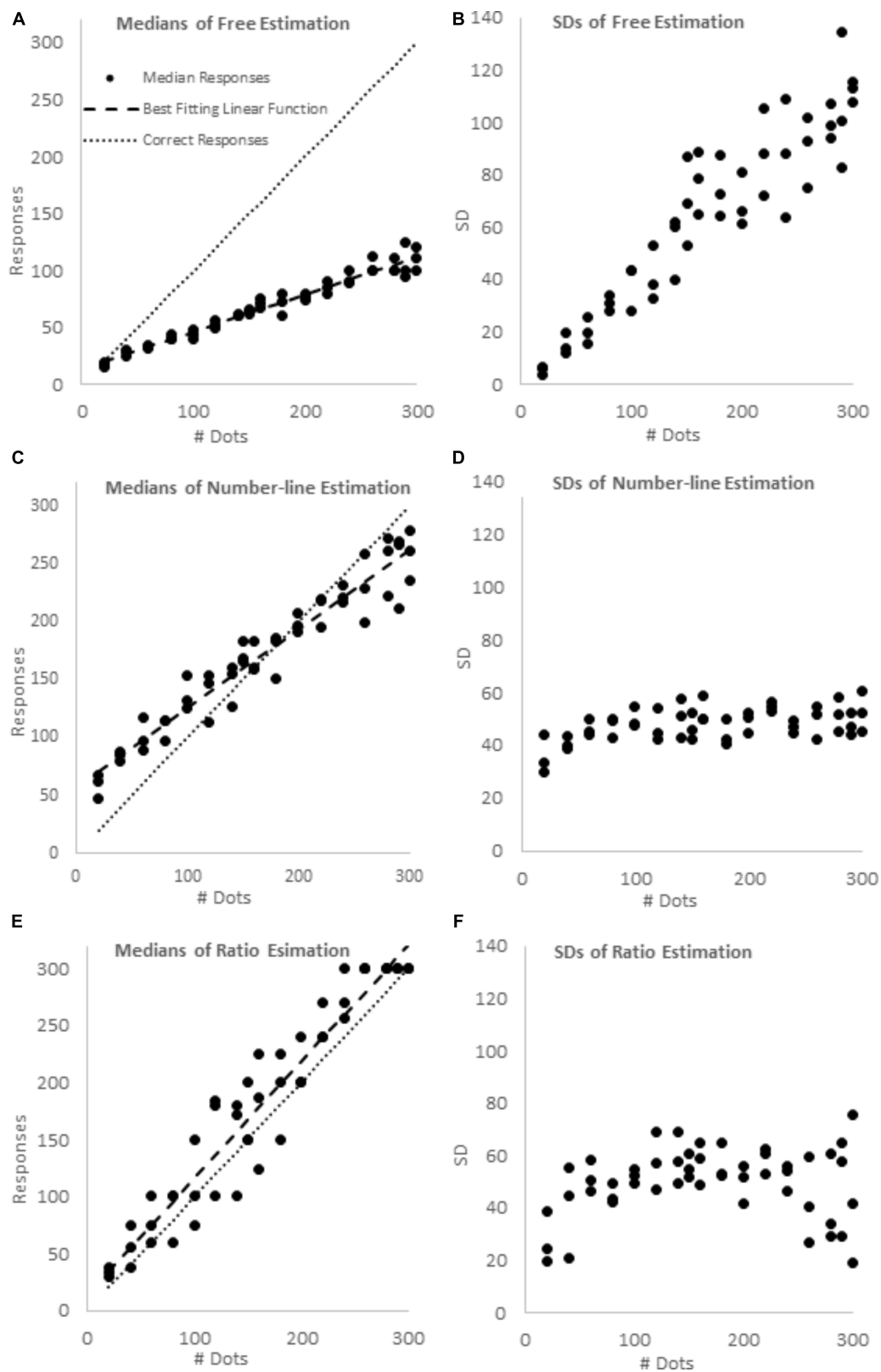


FIGURE 4 | Median responses and SDs in the free estimation (A,B), number-line estimation (C,D) and ratio estimation blocks (E,F).

responses, and the standard deviation of those responses. Plots of these data are presented in **Figure 4**. We fit the median responses to four different models:

$$\text{Linear: median} = B^* \text{stimulus} + C$$

$$\text{Logarithmic: median} = B^* \ln(\text{stimulus}) + C$$

$$\text{Power: median} = C^* \text{stimulus}^B$$

$$\text{One-cycle Cyclical Power Model: median} = (\text{stimulus}^B / (\text{stimulus}^B + (\text{Range} - \text{stimulus}^B))) * \text{Range}$$

For consistency, all models were fit by minimizing the sum of squares distance to the predicted value, and all R^2 s were calculated as $1 - (\text{Residual Sum of Squares})/(\text{Corrected Sum of Squares})$. Parameters B and C were allowed to vary freely in all models. The 1-cycle cyclical power model did not include a C parameter, but rather included a Range parameter, which indicates the range of values over which responses may be given. The 1-cycle model was run both with Range fixed at 300, and with Range allowed to vary, but constrained to be greater than or equal to the maximum median value in the data set. We utilized the nonlinear regression function of SPSS version 21 to conduct these analyses. A linear regression was also run on the standard deviations. Regression results are presented in **Table 1**.

Regressions

As predicted, only the free-estimation task showed scalar variability (see **Table 1** and **Figure 4**). Indeed, set size accounted for over 86% of the variance in SD for the free estimation task, but less than 22% of the variance in SD for the number-line task, and less than 2% of the variance in SD for the ratio estimation task. In the Number-line estimation trials, SD had little relationship with the stimulus, and in the ratio estimation trials, SDs appear lowest for the extreme proportions of 0 and 1, and to peak near 0.5.

Participants' median responses appeared to increase linearly with stimulus magnitudes in all three conditions (see **Table 1** and **Figure 4**). Indeed, for all three blocks, the linear model was a better fit than the logarithmic model and as good a fit as the standard power model. However, the ratio estimation and number-line estimation tasks were also well fit by cyclical power models, whereas a cyclical power model could not be fit to the free estimation task. Free estimation was the least accurate (Linear regression: slope = 0.327, intercept = 13.708), with responses consistently $\sim 1/3$ of the true value, and ratio estimation was the most accurate (Linear regression: slope = 1.020, intercept = 14.573), with responses quite near the true values. Number-line estimation had intermediate accuracy (Linear regression: slope = 0.683, intercept = 57.210). As would be predicted by a cyclical power model, median number-line estimates were overly high below the midpoint of the range, relatively accurate near the midpoint, and too low above the midpoint. We confirmed that this over- then underestimation pattern was significant using binomial tests.

TABLE 1 | Various regressions on median estimates and linear regressions on standard deviations for the free estimation, number-line estimation, and ratio estimation tasks.

Model	Free estimation	Number-line estimation	Ratio estimation
Linear			
B (SE)	0.327 (0.009)	0.683 (0.026)	1.020 (0.042)
C (SE)	13.708 (1.784)	57.210 (4.983)	14.573 (7.845)
R^2	0.961	0.933	0.926
Log			
B (SE)	36.607 (2.113)	77.081 (4.613)	113.414 (7.637)
C (SE)	-111.437 (10.497)	-207.444 (22.915)	-372.217 (37.937)
R^2	0.862	0.853	0.821
Power			
B (SE)	0.758 (0.030)	0.602 (0.030)	0.900 (0.051)
C (SE)	1.452 (0.231)	8.161 (1.289)	1.869 (0.512)
R^2	0.960	0.930	0.926
1-cycle cyclical power model, variable Range			
B (SE)	N/A	0.603 (0.028)	1.125 (0.108)
Range (SE)	N/A	368.188 (14.896)	300.000 (19.005)
R^2	N/A	0.934	0.891
1-Cycle cyclical power model, range fixed at 300			
B (SE)	N/A	0.581 (0.041)	1.125 (0.107)
R^2	N/A	0.879	0.891
Standard deviations			
B (SE)	0.364 (0.020)	0.033 (0.009)	0.018 (0.021)
C (SE)	2.765 (3.803)	42.384 (1.705)	46.496 (4.006)
R^2	0.869	0.212	0.015

For smaller arrays (i.e., 20, 40, 60, 80, and 100 dot arrays in each of the three formats) 15 out of 15 median estimates were greater than the stimulus values ($p < 0.001$). For larger arrays (i.e., 200, 220, 240, 260, 280, 290, and 300 dot arrays in each of the 3 formats) 20 out of 21 median estimates were less than the stimulus values ($p < 0.001$). However, the high and low endpoints failed to converge toward the anchors as we had predicted based on the cyclical power model.

DISCUSSION

Our results showed that task differences did in fact lead to vast differences in participants' abilities to make accurate estimates from ANS-processed stimuli. We found that free estimation yielded underestimates throughout the tested range. In contrast, number-line estimations first over- and then underestimated the size of the stimuli, though via a shallow linear slope as opposed to the predicted cyclical power model. Finally, performance on ratio estimation tasks was quite accurate. Indeed, ratio estimation yielded an unbiased linear map to symbolic number, whereas both the free and number-line estimation tasks yielded biased maps. Further, only the free estimation task exhibited scalar variability. These differences emerged even though all three tasks featured stimuli that current theory would suggest are processed by the ANS. Such results would not be expected

given the assumption that understanding of symbolic numbers is based on a direct mapping between number symbols and ANS-processed numerosities. These findings have implications for theories regarding the degree to which ANS-based estimation might serve as a good foundation for grounding symbolic number magnitudes.

Implications for Mapping Free Underestimation

Free estimations of dot arrays – a prototypical ANS task – led to considerable underestimates of the numerosities of the arrays, yielding the least accurate mappings of the three task formats. This is consistent with findings in prior literature (e.g., Indow and Ida, 1977; Izard and Dehaene, 2008; Crollen et al., 2011). Indeed, to our knowledge, free estimation of dot arrays has only proven accurate in three specialized situations: The first situation involves numbers in the subitizing range (up to ~4–5 objects), which recruits the object tracking system (e.g., Chesney and Haladjian, 2011). Second, free estimation for numerosities between 4 and 8 dots are also accurate on average, although estimates are less precise than in the subitizing range (e.g., Taves, 1941; Kaufman et al., 1949). In the third instance, some have found that free estimation, although not precise, is accurate on average, with larger arrays when feedback is given after every single trial to allow calibration (Minturn and Reese, 1951). However, Izard and Dehaene (2008) showed that this calibration can easily be thrown off by a single instance of inaccurate feedback.

This poses considerable difficulties for accounts that argue that the ANS-based magnitude perception serves as a ground for specific numbers. Given the failure of free estimation to facilitate accurate maps between numbers and their nonsymbolic analogs, it makes sense to question whether the ANS can be used to ground number symbols in a direct 1-to-1 fashion. For example, presuming that the ANS response to an array of 20 dots could serve as a stable referent for the symbol “20” seems untenable given the demonstrated inaccuracy of free estimation. This is not to say that we should abandon the ANS-as-ground position entirely. Rather, we believe it necessary to re-examine how ANS magnitudes and symbolic numbers might be linked. The current data may offer some insight into how this might be accomplished.

Performance on the free estimation task was very well fit ($R^2 = 0.961$) by a linear function with a slope of 0.327. Thus, although inaccurate, participants were quite reliable in their underestimation; they underestimated values at a consistent proportion of about 1/3. Of note, this particular underestimation yielded an estimate range with a maximum of approximately 100, even though the maximum array size was 300 dots. The large discrepancy is quite interesting, and we speculate that the value 100 may have a certain cultural status of being a default “large number.” This would explain why participants should happen to scale their responses so that the upper limit would be approximately 100. Given that prior research clearly demonstrates that adults can scale subsequent responses against a standard value (Izard and Dehaene, 2008; Thompson and Siegler,

2010), it is plausible that the 1/3 slope observed here was the result of “auto-scaling,” whereby participants assumed that the largest dot-set had 100 dots and scaled the remaining responses accordingly.

The Relational ANS

Although estimation patterns for all three tasks approximated linearity, ratio tasks clearly yielded the most accurate estimates. Median estimates were extremely well fit by a linear model with a slope of one and an intercept that was statistically equivalent to zero. Even the power model fit for ratio tasks yielded a Stevens' exponent of 0.9, indicating a curve that is very close linear. Considering this result in light of prior research showing that people can make proportion judgments cross-modally with great accuracy (Matthews and Chesney, 2015), this offers an intriguing possibility for grounding unfamiliar number symbols: Perhaps one way to gain an intuitive understanding for the magnitude of an unfamiliar number symbol is to start with a known number symbol and to use a cross-format proportion to convey how large the unfamiliar number is compared to the familiar number (see also Leibovich et al., 2016).

Chesney and Matthews (2013) found results consistent with this using number lines. They had undergraduates perform a number line estimation task using a line that extended from 0 to $0.999 \times 10^{4.5}$. Participants were unfamiliar with the magnitude of $0.999 \times 10^{4.5}$ (i.e., 31,591) and performed poorly until given the hint that 16,000 was roughly halfway along the number line. This intervention greatly improved performance. Participants used cross format proportion matching (Barth and Paladino, 2011; Sidney et al., 2017) to map the source ratio – the line segments' lengths – to the target ratio – the symbolic numbers. Thus they began to correctly treat $0.999 \times 10^{4.5}$ as roughly twice as large as 16,000, or about 32,000. The unfamiliar symbol gained meaning. A similar process can be used to map symbolic to nonsymbolic ratios more generally. For example, if a child watches her grandmother mapping 8 grapes to a “handful” in a recipe, and later saw 16 grapes being mapped to a “cup,” she could determine that the ratio of a “handful” to a “cup” was about 1:2, and use this knowledge in deciphering quantities in future recipes.

This process might be used by children learning symbolic numbers. If they observe a set of 25 dots being referred to as “20” and a set of 50 dots being referred to as “40” – such dot arrays are often underestimated (Taves, 1941; Izard and Dehaene, 2008; Crollen et al., 2011) and can even be purposefully mapped to larger or smaller values with inducers (Izard and Dehaene, 2008) – they can learn that the ratio of “20” to “40” is 1:2. The observed symbolic number to nonsymbolic numerosity map might be biased, but the nonsymbolic ratio is maintained. Such enumeration biases would be immaterial if relational mapping is the primary mechanism supporting the link between symbolic and nonsymbolic quantities. Moreover, if a system of ratios between symbols is known (e.g., “5” is half “10,” “10” is half “20,” “20” is half “40”), and at least one of the symbols is accurately mapped (e.g., five dots is

“5”) then a sense of scale for the other mapped symbols can propagate forward. Thus, it may be an approximate sense of proportion that drives the link between ANS estimation and symbolic number, rather than a direct correspondence between a symbolic number and a specific ANS magnitude. This perceptually based ratio sense would have limited utility compared to exact symbolic representations (e.g., one can symbolically represent 300/500 and 301/500, but one is unlikely to distinguish between their nonsymbolic instantiations) but all such perceptually based processes are necessarily limited in this sense.

Although this account is speculative, it is quite consistent with psychophysical accounts of how ANS-based comparison is processed. Indeed, as Sidney et al. (2017) observed, Weber’s law is fundamentally parameterized in terms of ratios, which means that existing conceptions of the ANS are largely compatible with viewing the system as inherently relational (cf., McCrink and Spelke, 2010; McCrink et al., 2013). This viewpoint essentially recapitulates Birnbaum and Veit’s (1974) observation that differences and ratios are in some sense mathematically equivalent in the logarithmically transformed space of perception, given that a log-transformed ratio yields a subtraction (i.e., $\log(x/y) = \log(x) - \log(y)$). We do note that work remains to be done to square this relational conception with neuroscientific evidence of numerosity specific neurons (e.g., Nieder et al., 2002; Diester and Nieder, 2007). That said, the mathematics of the dominant model is incontrovertible, so a relational conception of the ANS should not be easily dismissed.

The relational view of the ANS may suggest that two numerosities are better than one when it comes to facilitating maps to number symbols. Using two numerosities when mapping ANS magnitude to symbolic numbers solves a perennial problem with free estimations – specifically the vast individual differences in these estimates. Importantly, ratio perception establishes a correspondence among multiple instantiations of the same ratio, e.g., 10/15, 20/30, 50/75, etc. Thus, there is an inherent calibration for ratio judgments that may largely circumvent idiosyncratic scaling seen in single judgments. These observations converge with emerging theories about how ratio might be used to establish maps from perception of continuous magnitudes to specific numbers – as argued, for instance by Sidney et al.’s (2017) commentary on Leibovich et al.’s (2017) generalized magnitude system theory. They also converge with theories positing that ratio might be the preferred format for equating perceived magnitudes across different modalities (Balci and Gallistel, 2006; Bonn and Cantlon, 2017). All combined, we interpret the data as suggesting that the ANS is perhaps best understood as a system that perceives relations between numerosities, and as such may be more accurate when used to assess ratios as opposed to whole numbers. Future research should investigate this possibility.

Limitations and Future Directions

Memory Issues in Number-Line Estimation

As noted above, our prediction that performance in the number-line estimation task would be characterized by a cyclical power

model was not fully supported: although median estimates were overestimated below the midpoint of the range, relatively accurate near the midpoint, and underestimated above the midpoint, the high and low endpoints failed to converge toward the anchors. This may have been due to the speeded presentation protocol we used in order to ensure that participants could not count individual dots. As soon as the stimuli disappeared from view, they had to be maintained in memory and were thus subject to decay. Although this applies to all three tasks, this speed component may have specifically complicated the number-line task. Free estimation and ratio estimation tasks like those used here are typically conceived of as involving relatively direct estimation. However, the proportion judgment model conceived of by Spence (1990) and Hollands and Dyre (2000) involves explicit strategies whereby the observer pegs landmark values that result from segmenting the range (e.g., into halves or fourths) and subsequently estimates the remaining distance between the stimulus and the reference point. Memory decay may thus have more substantially impacted the bounded-estimation process than the other two tasks. In future work, we will compare performance in speeded and unspeeded conditions. We will also investigate potential differences in performance that might be induced by instructions focusing on an explicit ratio match versus instructions that focus on the landmark-based proportion judgment of the Spence (1990) model.

Free Estimation, Linear Compression

One interesting result specific to the free estimation task was that, although participants consistently underestimated the dot array magnitudes, their estimates did not appear compressive in the traditional sense that they were better fit by a logarithmic or power function than a linear function, or that the proportion of underestimation became greater as the set size increased. Rather, the portion of underestimation remained constant. This linear performance is more typical of sequentially presented stimuli than the simultaneous presentation we used here (Taves, 1941; Meck and Church, 1983; Cordes et al., 2001; Izard and Dehaene, 2008; Crollen et al., 2011). While this may have been an idiosyncrasy of our data set, it is possible that this was due to our choice of stimuli. Our smallest value, 20, was well above the subitizing range (~ 4 , Taves, 1941; Chesney and Haladjian, 2011). Numerosity estimates are known to be quite accurate when people subitize (Taves, 1941; Chesney and Haladjian, 2011). There also appears to be a benefit to accuracy when estimating values immediately above this range (e.g., 6, 7, 8; Taves, 1941; Kaufman et al., 1949), possibly due to subitizing based strategies: at the very least, these values would be known to be greater than ~ 4 . Our results show that people are linear with a slope less than 1 for larger values. Including both accurately assessed, subitizing-influenced low number values and underestimated higher values in a stimulus set would yield bi-linear performance. Regressions comparing compressive power or log functions to (mono-)linear functions for such bi-linear data would favor the compressive functions. Further work is needed to assess if (mono-) linear rather than compressive estimation patterns are typically seen when values that may be aided by subitizing strategies are excluded from consideration.

CONCLUSION

There are three main takeaways from these results. First, number-line estimation tasks appear to have limited utility in investigating either the ANS or the mapping between the ANS and symbolic numbers. These tasks do not yield the classic error patterns (i.e., scalar variability) seen in ANS estimation, and the functional form of performance on line-estimation tasks does not necessarily parallel the functional form of individuals' underlying magnitude representations. The use of nonsymbolic stimuli does not overcome these limitations. Second, the underestimation in the free-estimation task, particularly relative to the accurate performance on the proportion judgments task, is problematic for theories that propose a direct mapping between symbolic numbers and ANS estimation of specific nonsymbolic magnitudes. Third, we suggest that a system that uses a sense of ratio to link symbolic numbers to ANS-perceived magnitudes may overcome these difficulties. Future research is needed to address these possibilities.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations regarding human subjects research by the

Internal Review Board (IRB) of the University of Notre Dame. The protocol was approved by the IRB of the University of Notre Dame. All subjects gave written informed consent in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This research was made possible in part by support from the Moreau Academic Diversity Postdoctoral Fellowship Program of the University of Notre Dame.

ACKNOWLEDGMENTS

We would like to thank Michael Villano for his aid in creating the stimuli, and Nicole McNeil for her help and support.

REFERENCES

- Anobile, G., Cicchini, G. M., and Burr, D. C. (2012). Linear mapping of numbers onto space requires attention. *Cognition* 122, 454–459. doi: 10.1016/j.cognition.2011.11.006
- Balci, F., and Gallistel, C. R. (2006). Cross-domain transfer of quantitative discriminations: is it all a matter of proportion? *Psychon. Bull. Rev.* 13, 636–642. doi: 10.3758/BF03193974
- Banks, W. P., and Coleman, M. J. (1981). Two subjective scales of number. *Percept. Psychophys.* 29, 95–105. doi: 10.3758/BF03207272
- Barth, H. C., and Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Dev. Sci.* 14, 125–135. doi: 10.1111/j.1467-7687.2010.00962.x
- Birnbaum, M. H., and Veit, C. T. (1974). Scale convergence as a criterion for rescaling: Information integration with difference, ratio, and averaging tasks. *Percept. Psychophys.* 15, 7–15. doi: 10.3758/BF03205820
- Bonn, C. D., and Cantlon, J. F. (2017). Spontaneous, modality-general abstraction of a ratio scale. *Cognition* 169, 36–45. doi: 10.1016/j.cognition.2017.07.012
- Cedrus Corporation (2007). *Super Lab 4.5*. San Pedro: Cedrus Corporation.
- Chesney, D. L., and Haladjian, H. H. (2011). Evidence for a shared mechanism used in multiple-object tracking and subitizing. *Atten. Percept. Psychophys.* 73, 2457–2480. doi: 10.3758/s13414-011-0204-9
- Chesney, D. L., and Matthews, P. G. (2013). Knowledge on the line: Manipulating beliefs about the magnitudes of symbolic numbers affects the linearity of line estimation tasks. *Psychon. Bull. Rev.* 20, 1146–1153. doi: 10.3758/s13423-013-0446-8
- Cohen, D. J., and Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychon. Bull. Rev.* 18, 331–338. doi: 10.3758/s13423-011-0059-z
- Cordes, S., Gelman, R., Gallistel, C. R., and Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychon. Bull. Rev.* 8, 698–707. doi: 10.3758/BF03196206
- Crollen, V., Castronovo, J., and Seron, X. (2011). Under- and over-estimation. *Exp. Psychol.* 58, 39–49. doi: 10.1027/1618-3169/a000064
- De Smedt, B., Noël, M.-P., Gilmore, C., and Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends Neurosci. Educ.* 2, 48–55. doi: 10.1016/j.tine.2013.06.001
- Dehaene, S. (1992). Varieties of numerical abilities. *Cognition* 44, 1–42. doi: 10.1016/0010-0277(92)90049-N
- Dierker, L., and Nieder, A. (2007). Semantic associations between signs and numerical categories in the prefrontal cortex. *PLoS Biol.* 5:e294. doi: 10.1371/journal.pbio.0050294
- Fabbri, S., Caviola, S., Tang, J., Zorzi, M., and Butterworth, B. (2012). The role of numerosity in processing nonsymbolic proportions. *Q. J. Exp. Psychol.* 65, 2435–2446. doi: 10.1080/17470218.2012.694896
- Feigenson, L., Dehaene, S., and Spelke, E. (2004). Core systems of number. *Trends Cogn. Sci.* 8, 307–314. doi: 10.1016/j.tics.2004.05.002
- Gallistel, C. R., and Gelman, R. (2000). Non-verbal numerical cognition: from reals to integers. *Trends Cogn. Sci.* 4, 59–65. doi: 10.1016/S1364-6613(99)01424-2
- Hollands, J. G., and Dyre, B. P. (2000). Bias in proportion judgments: the cyclical power model. *Psychol. Rev.* 107, 500–524. doi: 10.1037/0033-295X.107.3.500
- Hollands, J. G., Tanaka, T., and Dyre, B. P. (2002). Understanding bias in proportion production. *J. Exp. Psychol.* 28, 563–574. doi: 10.1037/0096-1523.28.3.563
- Hurewitz, F., Gelman, R., and Schnitzer, B. (2006). Sometimes area counts more than number. *Proc. Natl. Acad. Sci. U.S.A.* 103, 19599–19604. doi: 10.1073/pnas.0609485103
- Indow, T., and Ida, M. (1977). Scaling of dot numerosity. *Atten. Percept. Psychophys.* 22, 265–276. doi: 10.3758/BF03199689
- Izard, V., and Dehaene, S. (2008). Calibrating the mental number line. *Cognition* 106, 1221–1247. doi: 10.1016/j.cognition.2007.06.004
- Jacob, S. N., and Nieder, A. (2009). Tuning to non-symbolic proportions in the human frontoparietal cortex. *Eur. J. Neurosci.* 30, 1432–1442. doi: 10.1111/j.1460-9568.2009.06932.x
- Jacob, S. N., Vallentin, D., and Nieder, A. (2012). Relating magnitudes: the brain's code for proportions. *Trends Cogn. Sci.* 16, 157–166. doi: 10.1016/j.tics.2012.02.002
- Kaufman, E. L., Lord, M. W., Reese, T. W., and Volkman, J. (1949). The discrimination of visual number. *Am. J. Psychol.* 62, 498–525. doi: 10.2307/1418556

- Kim, D., and Opfer, J. (2015). "Development of numerosity estimation: a linear to logarithmic shift?," in *Proceedings of the 37th Annual Meeting of the Cognitive Science Society*, Pasadena, CA.
- Krueger, L. (1984). Perceived numerosity: a comparison of magnitude production, magnitude estimation, and discrimination judgments. *Atten. Percept. Psychophys.* 35, 536–542. doi: 10.3758/BF03205949
- Leibovich, T., Kallai, A. Y., and Itamar, S. (2016). "What do we measure when we measure magnitudes?," in *Continuous Issues in Numerical Cognition*, ed. A. Henik (San Diego: Academic Press).
- Leibovich, T., Katzin, N., Harel, M., and Henik, A. (2017). From "sense of number" to "sense of magnitude": the role of continuous magnitudes in numerical cognition. *Behav. Brain Sci.* 40:e164. doi: 10.1017/S0140525X16000960
- Lewis, M. R., Matthews, P. G., and Hubbard, E. M. (2016). "Chapter 6 - neurocognitive architectures and the nonsymbolic foundations of fractions understanding," in *Development of Mathematical Cognition*, eds D. B. Berch, D. C. Geary, and K. M. Koepke (San Diego: Academic Press), 141–164. doi: 10.1016/B978-0-12-801871-2.00006-X
- Lyons, I. M., Ansari, D., and Beilock, S. L. (2012). Symbolic estrangement: evidence against a strong association between numerical symbols and the quantities they represent. *J. Exp. Psychol.* 141, 635–641. doi: 10.1037/a0027248
- Matthews, P. G., and Chesney, D. L. (2015). Fractions as percepts? Exploring cross-format distance effects for fractional magnitudes. *Cogn. Psychol.* 78, 28–56. doi: 10.1016/j.cogpsych.2015.01.006
- Matthews, P. G., and Lewis, M. R. (2016). Fractions we cannot ignore: the nonsymbolic ratio congruity effect. *Cogn. Sci.* 41, 1656–1674. doi: 10.1111/cogs.12419
- Matthews, P. G., Lewis, M. R., and Hubbard, E. M. (2016). Individual differences in nonsymbolic ratio processing predict symbolic math performance. *Psychol. Sci.* 27, 191–202. doi: 10.1177/0956797615617799
- McCrink, K., and Spelke, E. S. (2010). Core multiplication in childhood. *Cognition* 116, 204–216. doi: 10.1016/j.cognition.2010.05.003
- McCrink, K., Spelke, E. S., Dehaene, S., and Pica, P. (2013). Non-symbolic halving in an Amazonian indigene group. *Dev. Sci.* 16, 451–462. doi: 10.1111/desc.12037
- McCrink, K., and Wynn, K. (2007). Ratio abstraction by 6-Month-Old infants. *Psychol. Sci.* 18, 740–745. doi: 10.1111/j.1467-9280.2007.01969.x
- Mechner, F. (1958). Probability relations within response sequences under ratio reinforcement. *J. Exp. Anal. Behav.* 1, 109–121. doi: 10.1901/jeab.1958.1-109
- Meck, W. H., and Church, R. M. (1983). A mode control model of counting and timing processes. *J. Exp. Psychol.* 9, 320–334. doi: 10.1037/0097-7403.9.3.320
- Minturn, A. L., and Reese, T. W. (1951). The effect of differential reinforcement on the discrimination of visual number. *J. Psychol.* 31, 201–231. doi: 10.1080/00223980.1951.9712804
- Moyer, R. S., and Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature* 215, 1519–1520. doi: 10.1038/2151519a0
- Nieder, A., and Dehaene, S. (2009). Representation of number in the brain. *Annu. Rev. Neurosci.* 32, 185–208. doi: 10.1146/annurev.neuro.051508.135550
- Nieder, A., Freedman, D. J., and Miller, E. K. (2002). Representation of the quantity of visual items in the primate prefrontal cortex. *Science* 297, 1708–1711. doi: 10.1126/science.1072493
- Núñez, R. E. (2017). Is there really an evolved capacity for number? *Trends Cogn. Sci.* 21, 409–424. doi: 10.1016/j.tics.2017.03.005
- Opfer, J. E., Siegler, R. S., and Young, C. J. (2011). The powers of noise-fitting: reply to Barth and Paladino. *Dev. Sci.* 14, 1194–1204; discussion 1205–1206. doi: 10.1111/j.1467-7687.2011.01070.x
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends Cogn. Sci.* 14, 542–551. doi: 10.1016/j.tics.2010.09.008
- Revkin, S. K., Piazza, M., Izard, V., Cohen, L., and Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychol. Sci.* 19, 607–614. doi: 10.1111/j.1467-9280.2008.02130.x
- Reynvoet, B., and Sasanguie, D. (2016). The symbol grounding problem revisited: a thorough evaluation of the ANS mapping account and the proposal of an alternative account based on symbol-symbol associations. *Front. Psychol.* 7:1581. doi: 10.3389/fpsyg.2016.01581
- Sasanguie, D., and Reynvoet, B. (2013). Number comparison and number line estimation rely on different mechanisms. *Psychol. Belg.* 53, 17–35. doi: 10.5334/pb-53-4-17
- Schley, D. R., and Peters, E. (2014). Assessing "economic value": symbolic-number mappings predict risky and riskless valuations. *Psychol. Sci.* 25, 753–761. doi: 10.1177/0956797613515485
- Sidney, P. G., Thompson, C. A., Matthews, P. G., and Hubbard, E. M. (2017). From continuous magnitudes to symbolic numbers: the centrality of ratio. *Behav. Brain Sci.* 40:e190. doi: 10.1017/S0140525X16002284
- Siegler, R. S., and Opfer, J. E. (2003). The development of numerical estimation evidence for multiple representations of numerical quantity. *Psychol. Sci.* 14, 237–250. doi: 10.1111/1467-9280.02438
- Spence, I. (1990). Visual psychophysics of simple graphical elements. *J. Exp. Psychol.* 16, 683–692. doi: 10.1037/0096-1523.16.4.683
- Stevens, S. S. (1957). On the psychophysical law. *Psychol. Rev.* 64, 153–181. doi: 10.1037/h0046162
- Taves, E. H. (1941). *Two Mechanisms for the Perception of Visual Numerosity*. New York, NY: Archives of Psychology (Columbia University).
- Thompson, C. A., and Siegler, R. S. (2010). Linear numerical-magnitude representations aid children's memory for numbers. *Psychol. Sci.* 21, 1274–1281. doi: 10.1177/0956797610378309
- Valentin, D., and Nieder, A. (2008). Behavioral and prefrontal representation of spatial proportions in the monkey. *Curr. Biol.* 18, 1420–1425. doi: 10.1016/j.cub.2008.08.042
- Varey, C. A., Mellers, B. A., and Birnbaum, M. H. (1990). Judgments of proportions. *J. Exp. Psychol.* 16, 613–625. doi: 10.1037/0096-1523.16.3.613
- Whalen, J., Gallistel, C. R., and Gelman, R. (1999). Nonverbal counting in humans: the psychophysics of number representation. *Psychol. Sci.* 10, 130–137. doi: 10.1111/1467-9280.00120
- Whyte, J. C., and Bull, R. (2008). Number games, magnitude representation, and basic number skills in preschoolers. *Dev. Psychol.* 44, 588–596. doi: 10.1037/0012-1649.44.2.588
- Xu, F., Spelke, E. S., and Goddard, S. (2005). Number sense in human infants. *Dev. Sci.* 8, 88–101. doi: 10.1111/j.1467-7687.2005.00395.x

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Chesney and Matthews. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership